DOCUMENT RESUME

ED 357 603 FL 020 962

AUTHOR	Malcolm, Donald J.
TITLE	Test Development Manual.
INSTITUTION	Kuwait Univ., Safat. Language Center.
PUB DATE	[92]
NOTE	151p.
PUB TYPE	Guides - Non-Classroom Use (055)
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	Foreign Countries; Higher Education; *Language Tests; *Standards; *Test Construction; Test Content
IDENTIFIERS	*Kuwait

ABSTRACT

Various memoranda concerning language test development procedures and technical operations are compiled for staff at the Kuwait University Language Center from the Office of Tests and Measurement. The memoranda are of interest to Unit Test Representatives but also are intended to provide guidance to unit supervisors, course coordinators, and teachers in preparing statements of course objectives, formulating test-content specifications, writing and reviewing test items, and assembling final forms of tests. The document covers the following: job description of Unit Test Representatives, test production schedules, test development procedures, grammatical classification systems, test-form designation codes, behavioral objectives, item-writing mechanics, grammar/structure items, vocabulary items, reading comprehension items, items for measuring reading/listening skills, guide to reviewing multiple-choice questions, format, item-assembly procedures and form, reading/interpreting computer output, instructions for recordkeeping, calculating and using standard forms, staff assignments, and security systems for tests based upon the multiple-choice format and for item cards. (LB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED357603

KUWAIT UNIVERSITY
LANGUAGE CENTER
Tests & Measurement Office

TEST DEVELOPMENT MANUAL

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Victor W. Mason
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

by Dr. Donald J. Malcolm

FL020962

KUWAIT UNIVERSITY
LANGUAGE CENTER
Tests & Measurement Office

TEST DEVELOPMENT MANUAL

by Dr. Donald J. Malcolm

Table of Contents

Memorandum		Page
1.	Job Description, Unit Test Representatives	1
2.	Test Production Schedules	4
3.	Test Development Procedures for Final Tests	6
4.	Grammatical Classification Systems	12
5.	Test-Form Designation Codes	17
6.	Behavioural Objectives	20
7.	Test Content Specifications	29
8.	Item-Writing Mechanics	34
9.	Grammar/Structure Items	36
10.	Vocabulary Items	39
11.	Reading Comprehension Items	42
12.	Items Intended Primarily to Measure Either Listening or Reading Comprehension and those Intended to Measure a Complex of Integrated Skills	54
13.	Guide to Reviewing Multiple-Choice Questions	58
14.	Standardized Placement and Final Test Cover Pages	61
15.	Test Item-Assembly Form: Explanation of Procedures	63
16.	Test-Item Assembly Form: Sample Form	65
17.	Reading and Interpreting the Computer Output for Language Center Item and Test Analyses	67
18.	Instructions for Completing the "Summary Form: Item and Test Analysis Data"	73
19.	Summary Form, Item and Test Analysis Data	78

Memorandum		Page
20.	Item Cards for the Tests and Measurement File	81
21.	Item Accession-Number Table	85
22.	Code Designations for Various Item Types	88
23.	Test Cover Sheet	94
24.	Procedure for Calculating and Using Standard Scores	96
25.	Staff Assignments in Connection with the Administration of Placement Tests and the Processing of Answer Sheets	107
26.	Placement Test Tally Sheet	109
27.	Pre-Test Cover Page	117
28.	Purpose of the Norming Test at KULC	119
29.	Survey of the T&M Office	122
30.	Security Systems for Tests Based upon the Multiple-Choice Format and for the Item Cards	134

Preface

The Test Development Manual brings together under one cover the more significant memoranda of the Office of Tests and Measurement that pertain to test development procedures and to the technical operation of that office.

All of these memoranda are of direct concern to Unit Test Representatives. A number of them, however, are intended also to give guidance to unit supervisors, course coordinators and teachers in the preparation of statements of course objectives, in the formulation of test-content specifications, in the writing and reviewing of test items, and in the assembling of final forms of tests.

Undoubtedly, additional memoranda and revisions to prior memoranda will be forthcoming. For that reason, this manual has been assembled in loose-leaf form. Recipients are asked to keep their manuals up to date as new and revised materials are issued.

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-1 | مكتب الاختبارات والقياس

Date: 16th Sept., 1979 التاريخ :

Ref : الرقم :

MEMO

From Donald J. Malcolm

To Director, Unit Supervisors, Unit Test Representatives

Subject Job Description, Unit Test Representative (UTR)

1. General Nature of the Job

A Unit Test Representative is a teacher who serves as the liaison person between his or her unit and the Tests and Measurement Office. In general a UTR is responsible for facilitating the implementation of T&M Office programs, policies and procedures within his or her unit.

2. Appointment of UTRs and Tenure of Office

UTRs are to be appointed by their respective Unit Supervisors. Appointments are normally to be for one academic year and are to be made no later than the last day of May in the immediately preceding academic year. At this time Supervisors are to notify the Head of the T&M Office of their appointees. A UTR may be appointed for additional years of service if this is agreeable to the UTR and the Head of the T&M Office.

3. Qualifications for Appointment as UTR

Supervisors are to attempt to appoint as UTRs individuals who

a. are likely to stay on at the Language Center for some time

b. have had at least one course in testing as part of their professional preparation

c. have an interest in serving as UTRs

d. are respected by their colleagues and are capable of interacting diplomatically with them on controversial issues

4. Amounts of Time Expected of Individuals in Discharging their Duties and Responsibilities as UTR's

The amounts of time individuals will spend each week in connection with their work as UTRs will undoubtedly vary throughout the year as a consequence of varying demands in test production schedules. On the average, however, UTRs will be expected to spend 15-20 working hours per week on testing matters. Supervisors are, therefore, to adjust the teaching loads of UTRs accordingly.

5. The Accountability of UTRs

A. At the end of each month in the academic year, beginning with September and ending in May, each UTR is to prepare a report on his or her testing activities during that month. Copies of the report are to be sent to the Head of the T&M Office, to the UTR's Supervisor, and to the Office of the Language Center Dean.

B. At the end of the academic year, the Head of the T&M Office is to evaluate the performance of each UTR. Copies of these reports are to be sent to the Language Center Dean and to the Supervisor concerned.

6. Specific Duties and Responsibilities of the UTRs

A. To provide guidance to the professional staff in his or her unit in the preparation of course and unit objectives and in the derivation of test content specifications from those objectives.

B. To provide guidance to the professional staff in his or her unit in the writing of test questions and the assembly of these into test forms.

C. To secure the concurrence of the Unit Supervisor, Coordinators and teachers on test production schedules and to take steps necessary for subsequent adherence to the schedules.

D. To review the results of item analyses and to revise or reject questions with undesirable statistical characteristics.

E. To assemble forms of the placement test for preliminary administrations at the end of the fall term and final forms for administration in September of the next academic year.

F. To see to it that all testing materials (test books, answer sheets, instructions to proctors, etc.) needed for the September Placement Test administration are packaged and stored before the end of the spring semester.

G. To brief Unit Staff members on procedures and on their duties in connection with scheduled test administrations.

BEST COPY AVAILABLE

H. To discuss with Course Coordinators T&M office recommendations for changes in the drafts of final examinations.

I. To prepare, if their backgrounds in measurement and evaluations are adequate for the task and if they elect to do so, test analyses (i.e. critique based on the statistical characteristics of the tests administered within their units).

J. To assist in obtaining and compiling data that are needed from the Units in connection with special studies being undertaken by the T&M office in cooperation with the units.

K. To assist in implementing within the Units any measurement and evaluation procedures that have been recommended by the T&M office and endorsed by the L.C. Director.

L. To assist Unit Supervisors in the administration of unit testing activities.

M. To attend meetings of the UTRs that may be called by the Head of the T&M office.

N. To submit the monthly activity reports specified in paragraph 5A above.

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-2 | مكتب الاختبارات والقياس

Date: 6th November, 1979 التاريخ :

Ref الرقم :

MEMO

From Donald J Malcolm

To Unit Test Representatives

Subject Test Production Schedules

Early in each semester the T&M office will prepare and distribute to each UTR copies of a "preliminary" Test Production Schedule (TPS) for his or her unit. This schedule specifies the deadline dates by which the several steps in the production of placement and end-of-course tests are to be completed. A blank copy of the TPS form for the fall semester of 1979 is attached to this memo. In connection with the TPS, the UTR is expected to do the following:

1. Review the preliminary TPS with his or her supervisor and with the course coordinators to ascertain if the schedule is realistic from their points of view. If changes are recommended, the UTR is to notify the T&M office accordingly. The T&M office will then issue a revised TPS, if it is feasible to do so.

2. Monitor the test development activities within his or her unit to determine if the schedule is being adhered to. If there is slippage, the UTR is to notify the T&M office.

Adherence to the Test Production Schedules is important for two major reasons:

1. To avoid overburdening the AVA office with end-of-term requests for production of tests and related materials.

2. To avoid the crash assembly of final tests at the end of the term with inadequate or no reviews by unit teachers or by the T&M office.

The dates given in the TPS are final dates for completion of the work specified. If the work can be done prior to these dates, so much the better.

Test Production Schedule
Fall Semester 1979

Preliminary ☐
Revision ☐

T & M OFFICE
LANGUAGE CENTRE
KUWAIT UNIVERSITY

LC Unit _____ No. _____ Date _____

Week	Events	Placement Test	Final Examinations
Sept. 15-19	Classes begin, 15th		
Sept. 21-26			
Oct. 29-3			
Oct. 6-10			
Oct. 13-17			
Oct. 20-24			
Oct. 27-31	ID-AL-ADHA 10/29 - 11/2 Incl.		
Nov. 3-7	Midterms, 11/6		
Nov. 10-14			
Nov. 17-21	Al-Hijra 11/20		
Nov. 24-28			
Dec. 1-5			
Dec. 8-12			
Dec. 15-19			
Dec. 22-26	Xmas. 12/25		
Jan. 29-2	Last Day of Classes 12/31 New Year, 1/1		
Jan. 5-9	LC Finals 1/2 - 1/5 Incl		
Jan. 12-16			
Jan. 19-23	Spring Vacation 1/19 - 2/3 Incl.		

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-3 | مكتب الاختبارات والقياس

Date: February 6, 1980 التاريخ :

Ref : الرقم :

MEMO

From: Donald J. Malcolm

To: Director, Assist. Director, Unit Supervisors, Unit Representatives, Course Coordinators.

Subject: Test Development Procedures for Final Tests

NOTE: This memorandum supersedes the T&M Office memorandum of 22nd September 1979 on the same topic. Please update your copy of the Test Development Manual by replacing the obsolete version of TDM-3 with this revision.

Outline of Procedures

1. At the beginning of each semester, the T&M Office is to send to each ELU a Unit Test Development Production Schedule. This schedule specifies the deadline dates by which the Unit is to complete each of the several steps in the test development process.[1] Adherence to this schedule is essential to provide adequate time for item-writing, item and test reviews, and the printing of test materials.

[1] See TDM-2 in the Test Development Manual.

2. Early in the semester, the Unit Supervisor is to request each Course Coordinator to appoint a test development committee for his or her course. Each of these committees is to consist of the Coordinator, as Chairman, and two or more teachers of the course. In addition, the UTR is to serve as an ex officio member of each TD committee.

3. Early in the semester, the UTR is to meet with each Coordinator to prepare a production schedule for the final examination in that course. These schedules are to be compatible with the Unit Test Production Schedule mentioned above in paragraph (1). The UTR is to send a copy of each of the course schedules to the T&M Office.

4. As its first task, each TD Committee, with the technical assistance of the UTR, is to prepare a preliminary set of content specifications for its test.[1] These specifications are to reflect what the members of the committee consider to be the important instructional outcomes of the course.

5. Following completion of the preliminary content specifications for the final tests within his or her unit, the UTR is to meet with T&M Office personnel to review (1) the most recently administered finals for the courses under consideration, and (2) the preliminary specifications for the new tests. After this review, the T&M Office is to send a memorandum to each Course Coordinator, with copies to the Unit Supervisor and UTR, either

---

[1] The test content specifications serve as a blueprint for putting the test together. Among other things, the specifications state which language skills are to be assessed and what kinds of questions are to be used in the assessment. For more detailed information, see TDM memorandum # 7.

endorsing the specifications or recommending changes in them.

6. Test Development Committes whose specifications have been endorsed by the T&M Office or which accept the recommendations of that office are to proceed with the construction of their tests. If a committee finds that any of the recommendations of the T&M Office are inappropriate, the points of difference are to be resolved by the Course Coordinator in consultation with the T&M Office.

7. The Course Coordinator is to assign the item writing tasks (in accordance with the test content specifications) to as many of the teachers of the course as is feasible. By so doing, the work-load on any one teacher will be minimized and the talent brought to bear on the production of the new test will be maximized.

8. The UTR is to see to it that item writers are provided with copies of the TD Office memoranda that relate to the types of items they have been asked to prepare.

9. To facilitate the review of new items and their subsequent assembly into test forms, they are to be typed onto item cards in accordance with TDM-8. The UTR is to obtain these cards from the T&M Office and to instruct the teachers in his or her Unit in their use.

10. All newly-prepared items are to be reviewed critically by the Test Development Committee for the Course for which they are intended. The Committee is to be responsible for making decisions concerning the acceptance, rejection,

or modification of the items that have been submitted. The UTR is to be responsible for providing the committees with information pertaining to the mechanics and substance of review procedures. (See TDM-13.)

11. After all of the individual test items have been approved by the Committee, the initial type-written draft of the final form is to be prepared. This draft is to include the cover page, the instructions to students for responding to the various sections and parts of the test, and the answer sheet. The UTR is to be responsible for seeing to it that the initial draft conforms to T&M Office conventions with respect to matters of format.

12. Following completion of the initial draft of the final form, the Coordinator is to ask the teachers of the course who did not serve on the Test Development Committee to review the test individually and to submit to him or her their recommendations for changes. In addition the Supervisor of the Unit is to arrange to have <u>at least</u> two members of the instructional staff (preferably individuals <u>not</u> teaching the course in question) take the test as a student would take it. The copies of the test that are to be given to these reviewers are <u>not</u> to have the "correct" responses indicated on them. The reviewers, in addition to answering each question, are to enter on the pages of the test their comments pertaining to any difficulties they encountered in taking it.

13. The Test Development Committee is to revise its draft of the test in the light of the performance and comments of the reviewers. Copy for the final form is then to be prepared.

14. No later than the date specified in the Test Production Schedule, the Course Coordinator is to submit a copy of the test to the Unit Supervisor for his review. This copy is to be accompanied by the following documents.

(1) The Test Content Specifications

(2) The answer key for any multiple-choice questions in the test

(3) Instructions (rules and procedures) for reading and grading "supply-type" questions

While a Supervisor may be critical of individual items in the test, the main purpose of the review at the Supervisor's level is to assure compatibility between the measurement goals of the test (as inferred from its contents) and the overall instructional goals of the Unit. If the Supervisor is dissatisfied with the test, he is to return it to the Course Coordinator for revision. Otherwise, the supervisor is to forward the test copy to the T&M Office for review.

15. The T&M Office is to receive all final examinations of all required courses during the final week of November, in the case of the first semester, and during the final week of April, in the case of the second. Review of the tests by the T&M Office is to be completed by the 10th of the following month (December or May). The final examinations from all required courses of the same unit are to be submitted for review at the same time, to facilitate comparison of tests for length, level of difficulty, and appropriateness from the standpoint of skills development from one course to another.

16. Following this review by the T&M Office, the test development committees will make changes as mutually agreed upon. If, following these revisions, the Supervisor finds the test acceptable, he is to attach a cover sheet to the test copy and forward the test to the T&M Office. The cover sheet is to contain the following information:

1 Identification of the course for which the test is intended.

2. An estimate of the number of students who will take the test.

3. The number of different versions ("scrambles") of the test that will be required for the administration.

4. Identification of the test construction committee and the test reviewers.

5. A statement, signed by the Supervisor, to the effect that in his opinion the test constitutes an appropriate final test for the course for which it has been developed.

17. The final review by the T&M Office is to be concerned largely with matters of test format in the interest of facilitating scoring and subsequent item and test analyses procedures. Any needed changes in format will be worked out by the T&M Office in cooperation with the UTR and Course Coordinator.

18. The typing of final copy and the production of the required number of copies of the test and related materials either within the Unit or using the facilities of the AVA Office are to proceed as in the past.

DJM/st

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-4 | مكتب الاختبارات والقياس

Date : November 6, 1979 — التاريخ :

Ref — الرقم :

MEMO

From Donald J. Malcolm

To Unit Test Representatives

Subject Grammatical Classification System

The attached Grammatical Classification System is to be used in classifying each discrete-point grammar/structure item that is prepared for a final or placement test. The classification code is to be entered in the upper-left corner of the item card in the space provided.

GRAMMATICAL CLASSIFICATION SYSTEM

1. **Simple Parts of Speech**

A. **Verbs**

1. Copula (BE)

   a. affirmative (SV agreement)
   b. negative
   c. Wo or Aux. in Wh. or Yes/No questions
   d. tag questions
   e. short answer
   f. past tense
   g. dummy subject 1. there
      2. it

2. Present Tense

   a. affirmative
   b. negatives
   c. Wo or Aux. in wh or Yes/No questions
   d. tag questions
   e. short answer
   f. agreement with time

3. Past Tense

   a-f as above

4. Present Continuous

   a-f as above

5. Present Perfect

   a-f as above

6. Present Perfect Continuous

   a-f as above

7. "Going to" future

   a-f as above

8. "Will" future

   a-f as above

9.a. Passive - Present Tense

   a-e as above

9.b. Passive - Past Tense

   a-e as above

9.c. Passive - Present Perfect

a-e as above

9.d. Passive - Past Perfect

a-e as above

10. Past Perfect

a-e as above

11. Past Continuous

a-e as above

12. Future Continuous

a-e as above

13. "Used to" Past

a-e as above

14. Modals

15. "had better"

16. VP Complement

a. Ø + infinitive
b. to + infinitive

B. Nouns

1. mass or count nouns

a. agreement with verb
b. a little/a few
c. some/any
d. much/many

2. formation

C. Articles

1. definite/indefinite
2. demonstrative

D. Adjectives

1. formation (& irregular)
2. attributable position
3. predicate position
4. comparative
5. superlative
6. intensifiers (very & much)
7. equality

E. Adverbs

1. formation (& irregular)
2. comparative
3. superlative
4. intensifiers (very & much)
5. frequency

F. Question Words

G. Pronouns

II. Phrase Structures

A. Prepositional Phrases

1. time (in, on, at, for, since, until)
2. place (in, on, at)

B. Participial Phrases

1. present
2. past

C. Gerund Phrases

1. subject
2. complement

III. Clause Structures

A. Connectors

1. co-ordinating
2. subordinating

B. Parallelism

C. Adjectival

1. (who is sitting in the chair)
2. (which I saw (it) deleted object)
3. so ..... that

D. Adverbial

1. so ..... that
2. time

E. Result Clause

F. Noun Clause

G. Reported Speech

1. present
2. past

H. Reduced Relative Clause

I. If clause

1. past
2. future

Classification System prepared by
Jane Gaffney, UTR Commerce
Steven Spurling, UTR Arts
May 1979

DJM/st

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-5 | مكتب الاختبارات والقياس

Date: 18 October 1978 التاريخ :

Ref الرقم :

MEMO

From D J Malcolm

To

Subject Test Form Designation Codes

With the exception of informal classroom tests, each test constructed and administered within the Language Centre is to be identified by an alpha-numeric code. This requirement applies to placement, mid-term, and final tests; and to any other special-purpose tests that may be developed. The purpose of the code is to facilitate accurate record keeping of the Language Centre's testing activities.

The digits and letters in the code represent the following in the order given.

1. Language field

   A. Arabic
   B. English
   F. French
   R. Russian

2. Faculty

   A. Arts
   C. Commerce
   E. Engineering
   S. Science

3. Semester

1. Fall
2. Spring
3. Summer

4. Academic year

8. 1978-79
9. 1979-80
O. 1980-81
1 1981-82

Note: The academic year is considered to begin in Sept. The code number indicates the calendar year in which the academic year commences. A test form is to be coded according to the academic year in which it is given even though the calendar year may be the next higher digit. See examples. In the case of pretests or preliminary versions of a test (see 6), the academic year in which the final version is to be given is to be entered in the code no. even though the calendar year in which the preliminary is administered may be different. See example.

5. Type of Test

P. Placement
M. Mid-year
F. Final

6. Developmental Stage of Test

X. Pretest
Y. Preliminary
Z. Final version

Note: The above 6 items are to be used in all test form identification codes. The following 3 are to be used only if applicable.

7. Alternate (parallel) forms of same test

A. First form
B. Second form
C. etc

BEST COPY AVAILABLE

8. Scrambled Versions

1. First scramble (Base form)
2. Second scramble
3. etc

9. Course designation

If a test is intended for one course, the three digit number for that course is to be appended to the test code. If a test is intended for more than one course, the code numbers of the courses are to be appended in sequence with slant bars between courses.

Examples of Test Form Designation Codes

1. EC19PZ2

The second scramble of the final form of the ELU Commerce placement test for administration in the fall semester of the 1979-80 academic vear.

2. EC19PYA

Form A of a preliminary version of the ELU Commerce placement test for administration in the fall semester of the 1979-80 academic year.

Note: The administration date of the final form is to be given, not that of the preliminary form.

3. EE18FZ 121

The final test for ELU Engineering course 121 to be given in January 1979 (end of first semester of 1978-79 academic year).

4. EE28FZ 121

As in 3 above except that administration date is at end of spring term in the 1978-79 academic year.

4. EA1OMZ1 131

The midterm test, first scramble, to be given in course 13, ELU Arts, in the fall term of 1980-81.

~~4-6~~

Tests and Measurement Office
Language Center
Kuwait University

7th January, 1978

Memo:

From: Donald J. Malcolm

To: Supervisors, Course Coordinators
and ELU Testing Representatives

Subject: Behavioural Objectives

As a major step in the development of the new placement tests, each course coordinator is to prepare a statement of the "behavioural objectives" of the course for which he is responsible. These statements are absolutely necessary if the placement tests are to measure the various language skills, abilities and aspects of knowledge that need to be taken into account either in assigning a student to a course at the appropriate level or in exempting him from that course.

The most general purpose of any instructional course, including language courses, is to effect changes in the behaviour of the students who experience it. A change in behaviour has taken place when the individual can do something that he could not do previously or when he can do something at higher levels of complexity or excellence than he could previously. A statement, then, of the behavioural objectives for a particular course consists of an enumeration of the specific changes in student behaviours that the course is intended to produce. Such an enumeration is a necessary first-step in constructing achievement tests that are capable of assessing the degree to which students have acquired the stated competencies.

Attached to this memorandum for your guidance is a set of behavioural objectives that was developed through the joint efforts of the T&M Office and Mr. Edward Richards, the 099 course coordinator for ELU, Arts. Please note the following characteristics of this example:

1. Specific objectives are identified under six major headings. These are:

   A. Speaking abilities
   B. Listening abilities
   C. Reading abilities
   D. Writing abilities
   E. (Knowledge of) Structural points
   F. (Knowledge of) Vocabulary

The above classification of language skills and knowledge is a simplification of one that was formulated by John Carroll in connection with the design of a proficiency test in

BEST COPY AVAILABLE

English as a foreign language[1]. The four abilities are integrative in nature; the two areas of knowledge are discrete. This is reflected in the kinds of objectives that are listed under the several headings in the 099 outline. The ability of a person to function in each of the four major skill areas is conditioned by his knowledge of vocabulary and of grammatical structures. For this reason, the structures and vocabulary that the student is expected to know are defined separately and in detail in the accompanying set of behavioural objectives.

2. An integral part of the behavioural objectives concept is that the behaviours be observable. If they can be observed, they can be measured and evaluated. If they cannot be observed, they cannot be measured and evaluated. Perhaps the most difficult task involved in the preparation of a set of behavioural objectives is the reduction of an abstraction such as "understanding" or "appreciation" to observable actions. Abstractions of this sort are common in general statements of educational objectives.

Objective No. 1, under Listening Comprehension Abilities on page 2 of the attached paper, is of this type. It reads, "Ability to understand spoken standard English". Such a statement provides no insights into either what it is that the student is to understand (other than it is to be in the spoken form) or how the student who does understand is to be distinguished from the one who does not. The behaviours listed under "Operational definitions" of "to understand" make it explicit, in terms of observable student responses, what "to understand" means at the 099 level of instruction. Additionally, the "Linguistic boundaries" and "Situational boundaries" restrict the expected student responses to these contexts with which he can be expected to be familiar as a consequence of the course and other prior experiences.

It is to be noted that some of the objectives given for 099 are sufficiently specific as to require no further definition of terms. "The ability to ask the time and the ability to give the time when asked" is an example.

In the interest of facilitating the assembly of the placement tests---a process that will require the "integration" of the behavioural objectives for several sequential courses---the classification scheme described in paragraph No. 1 above should be used, to the extent possible, in stating

---

[1] Carroll, John. "Fundamental Considerations in Testing for English Language Proficiency of Foreign Students" in Allen & Campbell, Testing English as a Second Language, McGraw-Hill, 1972.

* The term "Operational definition" comes from the philosophy of science. It refers to the definition of a construct in terms that suggest how that construct is to be measured.

the instructional goals for each course. In some instances it may not be possible to classify neatly an expected language behaviour under one of the four primary abilities (for example, the ability of the student to take notes in English on material presented orally). In cases of this sort, additional major headings will be required in the outline. In addition, other major headings may have to be introduced to encompass objectives that are emphasized in the course but that are not uniquely associated with the acquisition of English as a foreign language (for example: the development of study skills).

It will also be noted that the 099 set of behavioural objectives is almost silent on the matter of "expected fluency". This is probably a reflection of the course level, a level at which fluency is not a primary concern. Coordinators of more advanced courses may wish to specify the expected fluency for various language behaviours. These may be incorporated into the outline by adding another sub-category following the specifications for "Linguistic boundaries" and "Situational boundaries".

The sets of behavioural objectives that you are requested to prepare will be more comprehensive in their coverage than will be the placement tests that are to be developed from them. In the placement tests, for example, it will not be possible---for logistical reasons---to assess speaking ability directly. However, other uses are to be made in the future of the outlines that are to be assembled now. One use is in connection with the preparation of end-of-course achievement tests. Another use is in connection with the development of testing instruments for research purposes.

Some concern has been expressed that putting the course objectives down on paper will tend to "freeze" the curriculum. This is certainly not an intent of the present exercise. If anything, the clear specification of what the various courses are attempting to do now---within units and across units---should faciliate orderly revision in the future.

Questions concerning the preparation of the statements of behavioural objectives should be directed to me at the T & M Office (514330).

Tests and Measurement Office
Language Center
Kuwait University

4th January, 1978

099 Arts, Behavioural Objectives

1. Speaking Abilities

Obj. 1. Ability to converse with speakers of standard English

A. Operational definitions

(1) "to converse"

a. to respond appropriately to discrete questions

b. to ask appropriate discrete questions

c. to engage in a brief but meaningful dialogue

(2) "Standard English"

a. English as taught in the course

b. English as spoken with a recognized educated accent

c. English at a normal rate of delivery

d. English with normal redundancy

B. Linguistic boundaries

(1) Structural:

Expected encoding and decoding operations within framework of structural and grammatical points identified in section V.

(2) Vocabulary:

Limited to that specified in section VI.

C. Situational boundaries
Content of conversational situations limited to:

(1) Material that is purely factual, informational, or descriptive.

(2) Material that is based on everyday activities and on events and places with which all students can reasonably be expected to be familiar (the weather, daily habits, life in Kuwait).

(3) Material that is familiar through prior class experiences (readings, dialogues, stories).

Obj. 2. Ability to ask the time and the ability to give the time when asked.

Obj. 3. Ability to describe orally objects and actions portrayed in a picture or series of pictures (Same linguistic and situational boundaries as described under 1 above).

Obj. 4. Ability to summarize orally material read in printed passages (Same linguistic and situational boundaries as described under 1 above).

II. Listening Comprehension Abilities

Obj. 1. Ability to understand spoken, standard English.

A. Operational definitions

(1) "to understand"

a. to carry out non-verbal actions in response to instructions or directions.

b. to paraphrase simple questions and single sentences presented orally.

c. to infer topic of two-or-three sentence dialogue

d. to infer location where two-or-three sentence dialogue takes place.

e. to answer questions based on the factual and informational content of a 150-word monologue.

f. to answer questions based on the factual and informational content of a dialogue of no more than three utterences with no more than two sentences in each utterance.

(2) "standard English":See I. 1A (2)

B. Linguistic boundaries: See I. 1B (1) & (2)

C. Situational boundaries: See I. 1C

III. Reading Abilities

Obj. 1. To understand communications written in standard English.

A. Operational definitions

(1) "to understand"

a. to carry out non-verbal actions in response to printed instructions or directions.

b. to select from a set of sentences the sentence that best answers a given question.

c. to select from a set of sentences the sentence that best paraphrases a given sentence.

d. to answer questions based on the factual and informational content of a reading passage no more than 150 words in length.

e. to identify the antecedents of pronouns used in a reading passage.

(2) "written standard English"

Simple or compound sentences that are grammatically acceptable pedagogically; Standard word order---
Noun phrase + verb phrase + adjective phrase.

B. Linguistic boundaries

(1) Structural and grammatical

Passages for reading to be based largely on points identified in section V. Other structures permissible when meaning can be inferred from context.

(2) Vocabulary:

Vocabulary in passage for reading to be limited mainly to that identified in section VI. Other high-frequency words used in class or in supplementary materials admissible. Some new words may be introduced if students can reasonably be expected to understand them.

C. Situational boundaries:

The content of printed communications to be limited to material that is purely factual, informational, or descriptive. Passages from newspapers, magazines, and books appropriate if edited for structural and lexical suitability. Material for eliciting "understanding" to be material that the student has not previously encountered or studied in the same form.

IV. Writing Abilities

Obj. 1. Ability to communicate in written English.

A. Operational definitions

(1) "to communicate"

a. to write simple sentences.

b. to write compound sentences consisting of two simple sentences joined by a high-frequency conjunction.

c. to write simple questions.

d. to write a number of simple sentences in sequence on one theme, given a series of leading questions.

C. Situational boundaries:

For evaluation purposes, writing limited to topics with which all students can be assumed to be familiar.

V. Structural Points

A high degree of proficiency is expected in ability to use correctly the following structures in speaking and writing, and to comprehend aural and written material based on them.

1. Verbs

A. Tenses of regular and irregular verbs in the time-referenced situations that are specified:

(1) Simple present and present continuous to express

a. timelessness (He goes to school on the bus).

b. limited duration (He is studying hard for the final examination).

c. the instantaneous (He is catching the ball).

(2) Simple past to describe an event that took place in the past and is regarded as completed (He wrecked his car.)

(3) "going to" future to express future or present intention (He is going to pay for the radio.)

2. Nouns, noun-substitutes

A. Countable and uncountable nouns

B. Determiners with countable and uncountable nouns

some/any much/many

C. Dummy subject "There"

(1) There is/are, was/were

(2) Distinguish from "They are"

3. Adverbs and adverbial phrases

A. Of frequency

(1) placement in sentence

(2) use of "ever" in questions

(3) specific adverbs of frequency

always never sometimes often

B. Of time

now yesterday tomorrow

morning/ afternoon/ evening/ night

last (night, week, month, year)

C. Of location

here there at(here)

in (the country) abroad

4. Adjectives

Limited to those that function as attributes and predicates.

No comparisons of adjectives.

5. Prepositions

Not isolated, but in prepositional phrases designating time and location (on Wednesday, on the table).

6. Sentences

A. Simple with correct word order

Sub + verb + obj + manner + place + time

B. Simple joined by

when until but because and

C. In the form of, " _____ want somebody to do something."

7. Questions (in tenses specified in I A above)

   A. Formation by

      1. inversions (using verbs "to be", "to have (got)", "must")

      2. Question tags (using operators is/are, do/does, did, must)

      3. "Wh" words ("when", "where", "who"---subj. and obj.---"how")

   B. Transformation of simple sentences to questions

8. Concord

   A. Agreement between subject and verb.

   B. Agreement in tense between question and answer.

VI. Vocabulary

By the end of the 099 course, students are expected to have as a minimum an active vocabulary consisting of the words listed in the back of Targets Books 1 and 2, plus the vocabulary of the supplementary material common to all 099 classes.

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office M-7 مكتب الاختبارات والقياس

Date : November 24, 1979 التاريخ : _____

Ref: الرقم : _____

MEMO

From D.J. Malcolm

To Unit Test Representatives

Subject Test Content Specifications

The content specifications for a test have been aptly described as the "blueprint" for the test. Ideally the specifications should be sufficiently detailed so that if two individuals, or committees, work independently in constructing forms of the test, the resulting products will be strictly parallel in terms of types of questions used, skills assessed, and level of difficulty. The two forms will be completely interchangable, and will differ only in their specific items.

Two terms are used in the material that follows that require definition. These are "Part" and "Section" as applied to the format of the test. "Part" refers to a major subdivision of the test. It is usually characterized by one or more of the following:

(1) a printed heading in the test book and/or answer sheet that designates it as a major skill area (Reading Comprehension, for example);

(2) a separate subscore that is entered in the scoring sheet or answer sheet; and

(3) separate timing, either actual or suggested. Parts are to be identified by title and number, as Part 1, Listening Comprehension.

"Section" refers to a subdivision of a part. It is usually characterized by (1) items of the same type,

BEST COPY AVAILABLE

and (2) special instructions to the students for responding to that item type. Sections are to be identified alphabetically, as Section A. In short, a test may consist of several major parts, each of which in turn consists of one or more sections.

A sample set of content specifications for a fictitious test is attached to this memo. In connection with these, please note the following:

1. The three parts of the specifications (Listening Comprehension, Reading Comprehension and Writing) provide examples, respectively, of specifications involving (a) discrete, multiple-choice items, (b) multiple-choice items based on passages, and (c) free-response questions.

2. The codes for the item types are given in the T&M Office Memorandum on Code Designations for Various Item Types (Memo No.22).

3. The total number of points for each section, as given in the content specifications, is the maximum that is possible under the scoring system used in reading the paper. If differential weights are used in connection with arriving at "marks", these should be given in the "Summary Table" at the end of the specifications.

Sample

Test Content Specifications

Name of Test Arts - Commerce Placement Test

Form EAC 11PZ Administration Date: Sept. 1981

Prepared by: _____

_____

Part 1. Listening Comprehension

Section A, Statements

(1)	Item type	LCS-4MC
(2)	Number of items	8
(3)	Number of points per item	1
(4)	Total number of points for section	8
(5)	Stimulus material	

Simple statements that sample vocabulary, structures and situations defined in behavioural objectives for Arts 099 & 131 and Commerce 099 and 101. One each appropriate for each course in each ELU.

Section B, Dialogues

(1)	Item type	LCD-4MC
(2)	Number of items	12
(3)	Number of points per item	1
(4)	Total number of points for section	12
(5)	Stimulus material	

Two-person dialogues, M&F voices.
3 each at each course level from 099 through 103 - 133. Vocabulary, structures and situations as defined in behavioural objectives.

Sample

Test Content Specifications

(continued)

Part 2, Reading Comprehension

		RCP-4MC
(1)	Item type	
(2)	Number of passages	8
(3)	Number of items	5 (average)
(4)	Number of points per item	1
(5)	Total number of points	40
(6)	Stimulus material	

For Arts: 1 passage not to exceed 150 words for each course: 099, 131, 132, & 133. Vocabulary, Structures and Situations as given in behavioural objectives. For Commerce: as above 099, 101 & 103

Part 3, Writing

Section A, Discrete Questions

		WSSNP
(1)	Item type	
(2)	Number of items	10
(3)	Number of points per item	2
(4)	Total number of points per section	20
(5)	Stimulus material	

Questions to elicit single-sentence replies without prompts. Cover range in Arts from 099 through 132, and in Commerce from 099 through 102. See behavioural objectives for relevant courses.

Sample

Test Content Specifications

(continued)

Part 3, Writing (continued)

Section B, Paragraph Writing

		WPNP
(1)	Item type	
(2)	Number of items	2
(3)	Number of points	10
(4)	Total number of points for section	20

(5) Stimulus material

Two assigned topics, each requiring a short paragraph in response. No prompts. Topics suitable for end of Arts 131 and Commerce 101

Summary Table

Part	Section	Time	No. of Points	Weight	No. of Marks
1. L.C.	A. Statements	30 minutes self-timed by tape	8	-	-
	B. Dialogues		12	-	-
	Sub - Total		20	2	40
2. R.C.		40 minutes timed	40	1	40
3. W.	A. Questions	50 minutes timed	20	-	-
	B. Paragraph		20	-	-
	Sub - Total		40	1	40
Total Test		120 minutes	100	-	120

KUWAIT UNIVERSITY

Language Centre

Date 21st March, 1978

Ref. ---

M-8

MEMO

Subject Some Item-writing Mechanics

From D.J. Malcolm

To Ms. Gaffney
Mr. Herbolich
Mr. Reimer
Mr. Spurling

1. Each discrete item is to be typed on a 12cm x 20cm (approximately 5" x 8") file card. Only one item is to appear on each card. Blank cards are available from the T&M office.

2. In the case of a set of items based on a passage, the passage is to be typed on one card; and a separate card is to be used for each item in the set (use a paper-clip to hold the set together).

3. Enter the key (correct response) on the back of the item card so that reviewers will not have access to the key when they read the item.

4. Enter in the upper-right corner of the card the ELU or ELU's for which the item has been developed (Arts, Commerce, Engineering, Science or in the case of items common to two ELUs Arts-Commerce, Engineering-Science). Also, if applicable, give the number(s) of the course(s) for which the item is appropriate. In the case of item sets, this information is required only on the first card in the set.

5. Enter, if applicable, in the upper-left corner of the card the following content information about the item:

   A. For all structure/grammar items, the structure/grammatical point tested by the item.

   B. For all reading comprehension items, the skill tested by the item.

   C. For all Engineering and Science items (in addition to A & B) the content classification of the item (description, observation, classification, definition or explanation).

6. The names of the item writer and of two reviewers are to be given on the back of the card.

7. A sample card is presented below:

Pronouns

Arts - Commerce
131 - 101

"I have forgotten my sandwiches."
"It doesn't matter, you can have _____."

A. some of us
B. some of ours
C. us some
D. some of our

FRONT

Key B

Item prepared by _____
Item reviewed by (1) _____
(2) _____

DJR/mm

BACK

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-9 | مكتب الاختبارات والقياس

Date: 14 October 1978 — التاريخ: _____

Ref: — الرقم: _____

MEMO

From D J Malcolm

To Supervisors, Teachers, Unit Test Representatives

Subject: Grammar/Structure Items

A grammar/structure section is to be included in each of the placement tests now being developed. This memorandum includes examples of some of the more commonly used item-types and gives some "pointers" for writing and reviewing such items. Definitions of terms used in connection with multiple-choice questions are given on the last page of this memo.

1. General specifications for grammar/structure items

   A. All items are to be of the multiple-choice variety with four options.

   B. Although examples of a variety of item types are given in this memo, each ELU is to select the two or three item types that appear to be most generally applicable to assessing the grammatical structures of concern to that Unit. The reason for this restriction is to minimize the special instructions that have to be given for taking the test.

   C. All stems (except in error-recognition items) are to be grammatical, depending on the nature of the structural point being tested.

2. Item types

   A. Sentence fill-in

      Ex. Karim _____ in Bahrain a few weeks ago.

      A. is arriving
      B. arrives
      C. has arrived
      D. arrived

BEST COPY AVAILABLE

Note: The above item type is used extensively because of the wide variety of structures that can be tested.

B. Dialogue fill-in

Ex. "We'll be ready to leave at six."

"Well, Jameela certainly _____ to be back by then."

A. must
B. ought
C. can
D. should

Note: the above type is particulafly useful in testing control of basic structures in spoken English.

C. Sentence Completion

Ex. Are these the students _____

A. you were talking about?
B. that you were talking about them?
C. about that you were talking?
D. you were talking about them?

D. Error recognition

Ex. The purpose of <u>the</u> (A) springs <u>are</u> (B) to hold the cover <u>down</u>. (C) <u>No error</u> (D)

Note: The student's task is to identify the underlined portion of the sentence that is ungrammatical. If he believes that the sentence is correct as it stands, he is to choose the "No error" option. In a set of such items, some grammatical sentences should be given. This item-type is used mainly in testing written expression.

E. Paraphrase

Ex. The solution has been in the tank for two hours.

A. The solution was putting two hours ago in the tank.
B. The solution was being in the tank for two hours.
C. The solution was put in the tank two hours ago.
D. The solution was into the tank two hours ago.

3. Item-writing pointers

A. Grammar items should be written so that only one option the key, is grammatical within the linguistic context of the stem. All other options should be ungrammatical when placed in the stem. These options should, however, be grammatical in and of themselves; eg when considered alone.

B. An effort should be made to control the semantic differences in grammar items. That is, where possible all distractors should hold meaning constant and vary only in grammatical form.

Good: The man _____ to the store yesterday.

A. is walking
B. walked
C. was walked
D. walks

Bad: The man _____ ___ to the store yesterday.

A. is jumping
B. walked
C. was talked
D. helps

C. Structure/grammar items are to test the student's control of elements of structure and grammar. Hence, the vocabulary used is to be as simple as possible. Similarly the reading load should be minimised. In other words, "the context should be at a lower level of linguistic ability than the actual problem being tested."

D. Only one grammatical or structural point is to be tested in any given item.

E. The stems are to be as short and as clear as possible. However, words which otherwise would have to be repeated in each option are to be included in the stem.

F. All distractors are to be reasonably attractive and plausible. They are to appear "right" to any person who doesn't know the correct answer. Suitable distractors can be derived from (a) mistakes in written work, (B) studen answers to "supply-type" questions, (C) teacher's experience, and (D) contrasting analyses between Arabic and English.

G. The key (correct response) is not to be conspicuously longer or shorter than the distracters.

H. In a series of items, the key should appear about an equal number of times in each of the possible choice positions (A,B,C, or D).

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-10 | مكتب الاختبارات والقياس

Date: 15th October, 1978 — التاريخ :

Ref : — الرقم :

MEMO

From D.J. Malcolm

To Supervisors, Teachers & Unit Test Representatives

Subject Vocabulary Items

A vocabulary section is to be included in each of the placement tests now being developed. This memorandum contains examples of appropriate item types and "pointers" for writing and reviewing such items. Definitions of terms used in connection with multiple-choice questions are given on the last page of this memo.

1. General specifications for vocabulary items

   A. All items are to be of the multiple-choice variety with four options.

   B. All items are to test vocabulary in context. However, definitions and examples may be used as stems. Discrete words or phrases followed by synonyms or paraphrases as options are not to be used.

2. Item types

   A. Multiple-choice fill-in

      The student is to select the word or phrase that logically completes the sentence.

      Ex. 1. The dog _____ at the stranger.

      A. roared
      B. barked
      C. squeaked
      D. pecked

Ex. 2. The specimen was dried in the _____ for twenty-four hours

A. desiccater
B. vibrator
C. perambulator
D. purifier

Another point to keep in mind when writing an item that contains an underlined word in the stem is that the options should either (depending on which is being tested) be at the same level of difficulty or at a higher level of difficulty than the individual word. Avoid having options that are "easier" than the underlined word.

Ex. 1 (Undesirable)

The thief was apprehended at the scene of the crime

(A) seen
(B) arrested
(C) caught
(D) hurt

Ex. 2 (Preferred)

The thief was caught at the scene of the crime.

(A) prosecuted
(B) apprehended
(C) victimized
(D) incarcerated

The reason for the above distinction is that items such as Ex 1 are more easily compromised in terms of test security than are items such as Ex 2.

The above item type is most appropriately used in those situations in which it is unlikely that the student will be familiar with synonyms for the key word.

B. Underlined words or phrases

The student is to choose from the options the word or phrase that is nearest in meaning to the underlined word or phrase in the stem. In practice, there seem to be two variations of this item type. In one, the distractors fit grammatically but illogically into the stem. In the second variation the distractors fit into the stem both grammatically and logically. However, only one option is synonymous with the underlined word in the context in which it appears. The second type is more difficult to write, but it is to be preferred because it prevents students from arriving at the correct answer by eliminating the distractors.

Ex 1. (Distractors fit grammatically but not logically)

Their success came about as a result of your valuable assistance

A. without
B. in spite of
C. because of
D. before

Ex. 2. (Distractors fit logically and grammatically)

He discovered a new route through the mountains

A. wanted
B. found
C. travelled
D. captured

C. Definitions

Ex: The instrument used to measure air pressure is called a

A. chronometer
B. hydrometer
C. barometer
D. hygrometer

D. Examples

Ex: Which of the following is a malleable substance?

A. Sugar
B. Hydrochloric acid
C. Rubber
D. Copper

3. Item-writing pointers

A. Vocubulary items should test vocabularly. Therefore they should avoid grammatical structures that students may find difficult to comprehend.

B. All options should be grammatically correct and should be grammatically correct when inserted in the stem.

C. All options should belong to the same word class as the word in the stem.

D. The correct option and the distractors should be at approximately the same level of difficulty. (If the correct option is more difficult than the distractors, the answer may be arrived at through elimination. If the correct option is easier, the better students may be misled.)

E. Avoid using pairs of opposites or synonyms as distractors.

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office M-11 مكتب الاختبارات والقياس

Date : 6th Oct., 1979 التاريخ : _____

Ref : Revision of Memo of Oct. 15, 1978 الرقم : _____

MEMO

From D.J. Malcolm

To Supervisors, Teachers and Unit Test Representatives

Subject Reading Comprehension Items

The memorandum contains suggestions for writing items (questions) to measure reading comprehension skills[1]. It is intended for staff members who will prepare or review reading comprehension items for the new placement tests. The suggestions, however, are also applicable to end-of-course tests.

Reading Comprehension Skills

To comprehend a written passage, a student must be able at least to (1) attach the correct meaning to a large percentage of the words in it, (2) cope with the grammatical forms it contains, and (3) respond correctly to the graphic symbols used[2]. At advanced levels of language proficiency, "verbal intelligence" is also an important component of comprehension. However, at the levels for which the placement tests are intended, the ability of the student to reason with words should not be a significant factor in determining his test score.

While knowledge of vocabulary, grammatical structures and printed symbols largely determine reading competence in the elementary and intermediate phases of language development, the testing of specific abilities in these areas can usually be done more efficiently through other types of items than those based on reading passages. Two exceptions to this generalization are (1) the ability to determine the meaning of unfamiliar words through contextual clues, and (2) the ability to identify the antecedents of words and phrases.

---

[1] Terms used in connection with questions are defined on page of this memorandum.

[2] Harris, David P. Testing English as a Second Language, McGraw-Hill, Inc. 1969 P 59.

Comprehension questions, depending on what the measurement goals are, can be based on written passages of virtually any length, from single sentences upwards. Most likely, reading comprehension material in the placement tests will vary from single sentences to paragraphs of not more than about 150 words.

Listed below are a number of comprehension abilities that can be tested through multiple-choice questions. For simplicity, the abilities have been stated in productive rather than in recognition terms. However, research indicates unambiguously that over a wide range of talent the ability to supply correct responses to questions correlates very highly with the ability to recognize correct responses to the same questions. Some examples of items are also given below, but the directions to the student have been omitted.

A. Comprehension abilities at the single-sentence level

1. The ability to paraphrase a given sentence:

   Ex: George has just returned from his vacation.

   A. George is spending his vacation at home.

   B. George has just ffnished his vacation.

   C. George is just about to begin his vacation.

   D. George has decided not take a vacation.

2. The ability to respond appropriately to a given question:

   Ex: "How does the amount of rain so far this year compare with that for the same period last year"?

   A. "During all of last year, two inches of rain fell".

   B. "The rain comes at the same time each year".

   C. "By this time, there was less rain last year".

   D. "The period will be the same this year as it was last year".

3. The ability to solve a simple problem based on the information contained in a given statement:

   Ex: The test was scheduled to begin at 8.00 AM, but it was delayed for fifteen minutes.

   At what time did the test start?

   A. Seven-forty-five in the evening
   B. Eight-fifteen in the evening
   C. Seven-forty-five in the morning
   D. Eight-fifteen in the morning

4. The ability to make a plausible inference from a given statement:

Ex: He bent over and dusted off his shoes.

A. His new shoes hurt his feet.

B. He was about to go in the water.

C. He had been walking along a sandy road.

D. He was very old.

B. Comments on comprehension items at the single-sentence level.

1. Advantages:

- The reading load per item is reduced to a minimum. Hence, a relatively large number of questions can be presented in a given amount of time. This contributes to test reliability.
- The sentences can be based on a wide variety of topics and can represent different styles of writing. This contributes to test validity.
- The single sentence has greater flexibility than the paragraph in testing comprehension problems associated with particular grammatical structures.

2. Disadvantages:

- The range of comprehension skills that can be tested is limited.
- The absence of contextual clues may make the comprehension of the isolated sentence more difficult than had the sentence been imbedded in a paragraph.
- In the absence of a context, the student may construe a meaning that is correct but at variance with the meaning intended by the writer of the item.

C. Comprehension abilities at the paragraph level.

Note: These abilities have been listed roughly in order of complexity.

1. The ability to answer informational questions which are answered directly in the passage in the same words as used in the question.

2. The ability to answer informational questions which are answered directly in the passage but in words other than those used in the question.

3. The ability to infer the meaning of unfamiliar words and phrases through the context of the passage.

4. The ability to identify the central thought in a passage in which the thought is explicitly stated.

5. The ability to identify the details of a passage that support the central thought.

6. The ability to identify the antecedents and/or referents of particular words or phrases in the passage.

7. The ability to identify the central thought in a passage in which that thought is not explicitly stated.

8. The ability to select an appropriate title for a passage.

9. The ability to derive inferences or to make deductions from the content of the passage. These may cover a variety of skills such as:

    a. Identifying the writer's purpose or point-of-view.

    b. Recognizing the "tone" of the passage.

    c. Using information contained in the passage to solve simple problems.

    d. Perceiving future implications.

D. Guidelines in selecting reading comprehension passages.

1. Passages on which sets of reading comprehension items are to be based should not exceed approximately 150 words. (The sample passage at the end of this memo contains 150 words.) Extended passages with a large number of questions based on each are to be avoided because certain topics typically give an undue advantage to certain students and an undue disadvantage to others. In the interest of a valid measure of reading ability, it is better to have several short paragraphs representing different topics and styles of writing than to have one extended paragraph of comparable length.

BEST COPY AVAILABLE

2. In general, paragraphs selected for reading comprehension are to be drawn from published sources (newspapers, periodicals, books) and are to be edited so as to be lexically and structurally appropriate for the course-level for which they are intended.

3. The passages selected are to represent good writing. They should be well organized, be logically sound, and be factually correct.

4. Passages should be of such a nature as permit a number of questions with unambiguous answers that cover the range of skills called for in the test specifications. In general, a passage of about 150 words should generate at least 5 meaningful questions. A passage of 150 words that yields but two questions is highly inefficient.

5. There are no fixed rules to serve as guides in choosing types of reading material that are likely to be fertile sources of questions for assessing a variety of reading comprehension skills. Passages of the following types, however, seem to lend themselves well to question formation:

   a. Those that deal chronologically with a series of events.

   b. Those that give instructions for doing something including any rules that are to be applied.

   c. Those that present different opinions on a given topic.

6. The subject matter of the passages is to be <u>new</u> to the students. At the easiest level, the content is to be a new configuration of the structures and vocabulary with which a student at that level can reasonably be expected to be familiar. At more advanced levels, the material dealt with should <u>not</u> have been encountered previously by the student. (The purpose of this requirement is to present, insofar as is possible to do so, the same task to each test taker.)

7. The level of difficulty of a passage intended to discriminate between students who are to be placed either in a lower-level or the next higher-level course covered by the placement test is to be such that students who have "passed" the lower-level course can handle the passage successfully but those who have failed that course cannot.

8. Lexical and structural elements in a passage are to be within the "linguistic boundaries" given in the test specifications for passages at a particular course level. Unfamiliar elements are admissible providing they are not crucial to understanding the passage or in answering specific items based on the passage. An exception to the latter point are questions that test the ability to determine meaning through context.

E. Types of passages to be avoided.

1. Passages that condense an extensive body of knowledge into a single paragraph

   Examples of these are 150 - 200 words summaries of the history of transportation or the descent of man. Such paragraphs, usually homemade, suffer from several defects. Among those defects are: (A) the sweeping generalizations are often over-statements of reality, (B) the students in varying degree already know something about the topic, (C) the generalizations do not lend themselves to a variety of meaningful questions, and (D) the questions can frequently be answered without reading the passage.

2. Passages that require "outside" information for their full comprehension

   Examples of these are (A) a technical passage in engineering or science that cannot be understood without prior knowledge of a physical law or theory, or (B) a description of an event that presupposes a knowledge of the geographical area in which the event occurs.

3. Passages that give an advantage to particular groups of students

   Examples of these are (A) passages based on topics usually covered in secondary school subjects that are taken by some students but not all, (B) passages based on topics relating to life experiences that some students but not all have had, (C) passages based on topics about which the members of one sex are likely to know more than the members of the other sex.

4. Passages that deal with information that is likely to be known by all students in the test population

   Items based on such passages are usually too easy, too trivial, or answerable without recourse to the passage.

5. Passages that have been taken directly out of the textbooks or other instructional materials used in specific courses

   Such passages encourage rote learning. Moreover, they give an advantage to those students who have taken the course in which the material was included.

F. Some guidelines for writing and reviewing questions (items) based on reading comprehension passages:

1. The questions should test significant material in the passages. Avoid items that are based on trivial detail.

2. The question and answer choices should be kept as simple as possible. Reading comprehension is to be tested through the passage, not through the wording of the questions.

3. Questions should not be answerable without reading the passage.

4. The stem of the item should either be in the form of a question or an incomplete statement which is sufficiently complete to make clear to the student what his task is. The stem should set the problem. The student should not have to read the answer choices (options) to determine the problem. The test of adequacy of an incomplete statement is whether sufficient information is given to permit the formulation of a meaningful question.

An inadequate stem	John	
	A.	C.
	B.	D.
A still inadequate stem	John went	
	A.	C.
	B.	D.
An adequate stem	John went to the store because	
	A.	C.
	B.	D.

5. Each item on a passage is to be independent of the other items in that passage. i.e. success in answering an item is not to depend on successfully answering a preceding question. Similarly, the wording of one question in a series is not to "give away" the answer to another question.

6. The measurement of reading comprehension skills should not be contaminated with the measurement of other language skills, such as writing ability. The use of "open-ended" or "supply-type" questions for ascertaining reading comprehension has at least two major drawbacks. First, the resulting score represents a mixture of reading and writing skills. In general, a student's reading ability will be underestimated if he cannot write well. Secondly, writing is a slow process. It is the slowest of the response modalities. Accordingly, the amount of reading material and the number of questions that can be given in a specified period of time is much less when students are required to write out answers rather than to select answers. The net effect of these two drawbacks is to diminish the validity of the test as a reading test and to reduce the reliability of the obtained scores.

7. The following points should be observed in preparing or reviewing the options (possible answers) for each item:

   A. There should be only one correct answer among the choices. Native speakers of English should have no difficulty in identifying the intended answer. To assure this, all questions and answer options should be "pre-tested" on native speakers before they are used in a final or placement test.

   B. Each option should be grammatically correct and should fit grammatically with the stem of the item. (Item writers tend to be conscientious in having the stem and the correct answer form a grammatically acceptable sentence. In the case of the other options, they sometimes overlook this point. Consequently, students may give the correct response not because they have understood the passage but because the correct response is the only one structurally acceptable).

   C. All distractors (wrong answers) should be reasonably plausible to students unsure of the correct answer. They should reflect the kinds of errors students make who do not

have the requisite comprehension skills for answering the question correctly. If the distractors are absurdities, students may arrive at the correct answer by a process of elimination.

D. The right answer (the key) should not be longer than the incorrect answers (the distractors). Sometimes item writers will add restrictive clauses to the key in order to make it unambiguously correct. Test-wise students soon learn that the longest option is most likely the correct one. In the interest of parsimony, all options should be as short as possible. If, however, the key has to be relatively lengthy, one of the distractors should be made about the same length to counterbalance it.

E. As is the case with all multiple-choice questions, avoid using "specific determiners" in the options. A specific determiner is a word such as "never" or "always". Students soon learn that options containing such words are usually incorrect.

F. As is the case with all multiple-choice questions, avoid pairs of options that are direct opposites in meaning. Such pairs provide the student with clues to the correct response. Either one of the pair is correct (which is usually the case) or both are incorrect.

G. Overlapping options should not be used, otherwise a good student may select the first option that is reasonably correct and may not even bother to read on to find that subsequent options are even more correct. For example, a passage might state that Dave's plane was due in at 3:00 pm but that it was delayed more than 3 hours. A poor comprehension question with overlapping options is the following:

Q. What time did Dave's plane arrive?

A. After 1:00 O'clock

B. After 3:00 O'clock

C. After 6:00 O'clock

Obviously, all of the options are correct.

H. Also to be avoided are options that are partially correct and partially incorrect. For example, a passage might state that Ibrahim went to the market to purchase some lamb, eggs and melons. A poor comprehension question with partially correct option is the following

Q. What did Ibrahim purchase at the market?

A. lamb, chicken, and fruit

B. meat, eggs and vegetables

C. etc.

D. etc.

Sample: Reading Comprehension Passage

The world's mineral wealth is running out. Even the United States, one of the richest natural storehouses of minerals on earth, is facing shortages. Between 1900 and 1950, when its population doubled, its consumption of minerals went up by six times, and by 1975 that rate will probably double. Already, world reserves of such essentials as copper, lead and zinc are considered inadequate. As the search for new sources goes on, it is clear that methods of exploration and ore extraction must be vastly improved to keep costs from soaring. But as high-quality ore in Minnesota neared exhaustion a few years ago, production of iron from taconite, a hard rock containing ore, was achieved. What happens when all the iron in the world is used up? Substitutes may be found, or perhaps man will be able to fish nickle-iron asteroids from space to supply earth with ore.

1. The main topic of this passage is

   A. the mineral wealth of the United States.
   B. new materials to replace metals.
   C. mineral shortages in the world.
   D. making iron from taconite.

2. About how many times greater was the consumption of minerals in the United States in 1975 than in 1900?

   A. Two.
   B. Six.
   C. Twelve.
   D. Twenty-four.

3. The paragraph mentions the rate of increase in the United States of all the following EXCEPT

   A. the population between 1900 and 1950.
   B. the use of minerals between 1900 and 1950.
   C. the population between 1950 and 1975.
   D. the use of minerals between 1950 and 1975.

4. When the writer of the passage refers to the United States as "one of the richest natural storehouses of minerals in the world" (line 2), he means that the United States has

   A. more money than most other countries.
   B. large supplies of minerals that are stored in special buildings for future use.
   C. more minerals in the earth than most other countries.
   D. larger supplies of minerals than it can possibly use.

5. To which of the following does "the search for new sources" refer (line 8)?

   1) Locating additional supplies of minerals on earth.
   2) Finding cheap ways to use low-quality ores.
   3) Finding materials to use in place of copper, lead and zinc.
   4) Locating iron on asteroids close to the earth.

   A. 1 and 2 only.
   B. 3 and 4 only.
   C. 1, 2 and 3 only.
   D. 1, 2, 3 and 4.

6. The very high prices that concern the writer of the paragraph are the prices in the future for

   A. metals.
   B. exploration,
   C. ore extraction.
   D. development of substitutes.

7. The production of iron from taconite is an example of

   A. the benefits that come from exploration.
   B. the increasing demand for iron in the modern world.
   C. a new method for getting iron ore.
   D. the use of a substitute for iron ore.

8. Which one of the following is the most appropriate title for this passage?

   A. The Earth's Minerals.
   B. Metals from Minerals.
   C. Iron in Modern Technology.
   D. The Irreplaceable Ores.

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-12 | مكتب الاختبارات والقياس

Date : November 14, 1979 التاريخ :

Ref : الرقم :

MEMO

From D.J. Malcolm

To Supervisors, Course-Coordinators & Teachers

Subject Items that are intended primarily to measure either listening or reading comprehension skills and those that are intended to measure a complex of "integrated" skills.

Listening Comprehension items at the level with which the Language Centre is generally concerned consist typically of a taped passage or dialogue followed by a series of printed questions to ascertain whether the student has understood the oral message. Depending on the particular test, the student is asked to indicate his answers in one of several possible ways---by selecting one of several printed options provided (multiple-choice), by supplying a verbal answer in written form (supply-type), or by giving a non-verbal response (such as performing certain operations on a map or table). Regardless, however, of the response format that is used, if the listening test is to be primarily concerned with assessing the ability of students to respond effectively to the phonology, stress and intonation of the spoken language, care should be taken to assure that as far as possible the scores that result from the test reflect listening comprehension skills rather than reading or writing skills or knowledge of vocabulary and structure. Some implications of this principle for the selection of appropriate oral passages or dialogues and for the preparation of the questions based on them are these:

1. Grammatical structures and vocabulary used in the stimulus material should be familiar to the examinees as a consequence of their prior course experiences. If new structures or vocabulary appears in the passage, comprehension should not be dependent upon them.

2. Printed questions based on the oral material (and the printed options to multiple-choice questions) should be well within the reading comprehension levels of the students tested.

3. Supply type questions based on the oral material should require responses that are as brief as possible and that are well within the writing ability levels of the students tested.

4. If non-verbal responses are used, care should be taken to assure that the directions for making the responses and the procedures for following them be as simple as possible.

A good pretest of a new listening comprehension test consists of administering it as a reading comprehension test to a representative sample of the kinds of students for whom the test is intended. If most of these students do not get nearly perfect scores on it, then successful performance on the test is obviously being affected by skills other than those associated with listening comprehension; and the test should be revised.

Reading comprehension items at the Language Centre usually consist of a passage followed by a series of questions in either a multiple-choice or a supply-type format. If the test is to be concerned primarily with ascertaining the ability of students to comprehend printed material at the level of difficulty of the passages, then the problems of comprehension should lie within the passages rather than within the questions based upon them. An implication of this principle for the test construction is that the questions (and options in a multiple-choice format) should be as simple as possible, requiring reading comprehension skills at a lower level than those required for comprehension of the passages. In the case

of supply type questions, the written responses that are called for should be as simple as possible and be well within the writing competencies of the students who are to take the test. Stated otherwise, the written responses that the examinee is called upon to make should be the minimal responses needed to indicate comprehension, or the lack thereof. If extended written responses are required, the resulting test scores will not reflect reading comprehension scores per se but a combination of reading and writing skills. Under certain circumstances, such composite, or "integrated" scores may be regarded as desirable. This matter is discussed in the last part of this memo. Under other circumstances, such as are considered in the next paragraph, they may not be desirable.

One type of item that has been used in some of the LC final examinations in courses in which the improvement of both reading and writing skills are instructional objectives is characterized by (1) a reading passage suitable for testing reading comprehension at the end of the course for which the final is intended and (2) a series of supply-type questions relating to the passage, directing the student to write his answers in complete sentences.(Prompts may or may not be given.) The student's score on the item is thus a function of the two skills. If he does not comprehend the passage correctly, his writing performance will be impeded and his writing abilities will likely be underestimated. If he comprehends the passage correctly but does not write well, his reading ability will be underestimated. In short, double-barreled items of this sort expose the student to "double-jeopardy". If the goals of a course are to develop reading skills and writing skills, they should---insofar as possible ---be tested separately. The type of item described at the beginning of this paragraph can be written in such a way that it puts the emphasis either on the writing or on the reading (but not both). If the emphasis is to be on writing and the resulting score is to be considered as indicative of writing ability, then the passage should be at a low enough level of difficulty so that it can be assumed that virtually all students for whom the test is intended can readily comprehend it. Only by so doing is it reasonable to assume that

all students are confronted with the same task in preparing their responses to the open-ended questions. If, on the other hand, the primary purpose of the item is to test reading comprehension, then the passage should be at a level of difficulty appropriate for that purpose, but the written responses required of the student should be kept as simple and short as possible. Moreover, the responses should be scored entirely on the basis of content rather than on the basis of correctness of expression.

In some instances an "integrated" test of reading and writing skills or of listening and writing skills is justifiable, if the course objectives aim directly at the developing of skills involving the processing of input information and expressing the results of the processing in written form. An example of an integrated reading-writing test is the preparation of a short, written summary of the material contained in a reading passage. An example of an integrated listening-writing test is the writing of a short memorandum based on a telephone communication. In order to prevent tests of these types from being primarily tests of a single component, i.e reading comprehension, listening comprehension, or writing ability, the stimulus material should be well within the grasp of most students for whom the test is intended; and the written responses that are called for should involve only elements of vocabulary and structure with which most of the students are familiar. In short, the test should place emphasis on the integration of the basic skills rather than on the basic skills themselves. The latter can be better assessed by other types of items.

DJM/ST

KUWAIT UNIVERSITY
Language Centre

Date 15th October, 1978

Ref.

M-13

MEMO

Subject Guide to Reviewing Multiple-Choice Items

From D.J. Malcolm

To Supervisors
Teachers
Unit Test Representatives

Each item written within an ELU for its new placement test is to be reviewed independently and critically by two teachers in that Unit other than the writer of the item. The item review assignments and the administrative arrangements for conducting the reviews should be made by the Supervisor in consultation with the Unit Test Representative. Reviewers may include the Unit Test Representatives, Course Coordinators, and other teachers.

Reviewers should read the "Pointers for Item Writing" that are contained in the T&M memos on the writing of reading-comprehension, vocabulary and grammar/structure items. They should also familiarize themselves with the "guide" that is given below and make use of it during the review process.

The items that are to be reviewed by a particular person will be given to him or her by the Unit Test Representative. Each item will be on a 5"x8" card. In the case of reading-comprehension and structure/grammar items, the skill or structural point that the item is intended to measure will be indicated in the upper-left corner of the card. The course for which the item is considered to be appropriate will be designated by number in the upper-right corner. The key (answer) will be on the back of the card. Reviewers are to look at the key only after they have arrived at their own answer to the problem.

If a reviewer is satisfied with an item as it stands, he or she is to sign the back of the item card. If a reviewer is dissatisfied with an item, he or she is to write his suggested changes (which may include scrapping the item) on a separate piece of paperand clip it to the item card. The item cards are to be returned to the Unit Test Representative. The revised cards will subsequently be sent back to the reviewer for approval.

Guide to Reviewing Multiple-Choice Items[1,2]

A. With respect to the item as a whole, consider whether

1. it tests effectively the skill that it is intended to assess. (Good vocabulary items test vocabulary, not control of structures. Good structure items test structure, not vocabulary. Good reading comprehension items require that the passages on which they are based be read and comprehended in order to arrive at the correct responses);

2. there is a significantly better way to test what the item tests. (A different type of item than that used might be better suited for testing the language problem);

3. it is within the appropriate range of difficulty for the intended test population. (Although the typical test should contain items of a broad range of difficulty, those that are answered by less than 10 percent or more than 90 percent of the population do not contribute significantly to valid measurement).

B. With respect to the stem, consider whether it

1. poses for the examinee a clearly defined problem or task. (The stem should make it clear to the student what the problem is. He should not have to read the options to determine his task. Stems consisting of a single word are almost invariably undesirable);

2. contains unnecessary information (only that information should be presented in the stem that is necessary for answering the item);

3. can be worded more clearly or concisely. (The principle of parsimony should be adhered to rigorously by item writers).

C. With respect to the options, consider whether

1. they fit logically and grammatically with the stem (application of this point will vary with

---

1. Adapted from "Manual of Information for Test Committee Members", Test Development Division, Educational Testing Service, Princeton, N.J.

2. Terms used in connection with multiple-choice questions are defined on the last page of this memorandum.

BEST COPY AVAILABLE

the basic skill being tested. In reading comprehension items, all options should fit the stem logically and grammatically. In vocabulary items, all options should fit the stem grammatically but may or may not - depending on the item type - fit logically. In structure/grammar items, the options may fit the stem neither structurally nor logically);

2. they can be worded more clearly or concisely (Words repeated in each option can often be incorporated in the stem).

D. With respect to the key (the correct or best answer), determine

1. which option you think is intended to be the correct response. (Try to answer the item <u>before</u> you refer to the key on the back of the item card);

2. whether the key actually answers the question posed (This point is particularly relevant to the review of reading comprehension items. A related point is whether the keyed answer is obtainable either directly or indirectly from the passage to which it refers);

3. whether the key needs to be made less obvious relative to other options or the stem (Should it be made longer? Shorter? More detailed? More abstract?).

E. With respect to the <u>distracters</u>, consider whether

1. there is any possible justification for considering one of them an acceptable response to the question (There should be <u>one</u> and <u>only one</u> acceptable and defensible answer to a given item. Errors commonly made by native speakers of English should not be used as distracters);

2. they are sufficiently plausible to attract examinees who are misinformed or inadequately prepared. (The distracters must not be a set of irrelevant responses).

DJM/ma

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-14 | مكتب الاختبارات والقياس

Date : December 1st, 1979 التاريخ : _____

Ref : الرقم : _____

MEMO

From Donald J. Malcolm

To Unit Test Representatives

Subject Standardized Cover Page for Language Center Placement Tests

The T&M Office recommends that the attached cover page be used, with modifications as needed, on all Language Center placement tests.

DO NOT OPEN THIS TEST BOOKLET UNTIL YOU ARE TOLD TO DO SO

EAC 10PYB

SN:_____

ENGLISH LANGUAGE TEST
Language Center
English Language Units
Colleges of Arts & Commerce
Kuwait University

Directions:

تعليمـــــات :

1. Write all your answers in pencil on the special Answer Sheet. Do not write anything in this test booklet.

١) اكتب جميع اجاباتك في ورقة الاجابة المخصصه لذلك . اكتب اجـابتك بالقلم الرصاص . * لا تكتب اى كلمة او علامة على هذا الكتيب

Each question in this test is followed by four answers. Only one answer is correct. Put an X in the box on your answer sheet to show the answer you think is correct. Make your X very dark and easy to see.

* هناك اربع اجابات لكل سؤال فـي هذا الاختبار منها واحدة صحيحه فقط ضع علامة x في المربع المناسب على ورقة الاجابة .

If you are not sure of the correct answer to a question, make the best guess that you can, and go on to the next question. Erase any answer you wish to change.

* اذا كنت غير واثق من الاجابة الصحيحه فعليك ان تحاول جهدك في استنتاجها او تخمينها ثم انتقل الى السؤال التالى .

2. This examination consists of three parts. It is timed as follows:

٢) يتكون هذا الاختبار من ثلاثة اجزاء مقسمه زمنيا كما يلى :

PART	TIME	
1. Grammar	12 minutes	— القواعد ١٢ دقيقـه
2. Vocabulary	13 minutes	— المفردات ١٣ "
3. Reading Comprehension	30 minutes	— القراءة للفهم ٣٠ "

3. If you finish one part of the examination early, Do Not go back to another part or begin a new part. Wait until the examiner tells you when to begin a part and when to stop.

٣) اذا انتهيت من جزء لا تبدأ فـي الجزء التالى الا بعد ان يعلن المتحن التعليمات بالبدء فـي الجزء التالى لجميع المتقدمين .

62

DO NOT OPEN THIS TEST BOOKLET UNTIL YOU ARE TOLD TO DO SO
لا تفتح هذا الكتيب الا بعد سماع الاشارة بالبدء

62

BEST COPY AVAILABLE

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-15 | مكتب الاختبارات والقياس

Date: 17th October, 1979 — التاريخ :

Ref: — الرقم :

MEMO

From: Donald J. Malcolm

To: UTRs

Subject: Test Item-Assembly Form: Explanation of Procedures

1. The Test Item-Assembly Form (see sample, following) is a record of the composition of the items making up a given test, as well as an indication of the degree of relative effectiveness of each item on that test in ranking the subjects who took it. The TIAF, as revised, provides a cross-reference with the item-card file (see M-20) for the efficient retrieval, via the accession-number system, of any item suitable for inclusion in the assembly of a later test form. The accession number is a hyphenated, seven-digit identification code located in the first column of the revised Test Item-Assembly Form (TIAF) and in the upper right-hand corner of each item card.

2. At the top of the TIAF, information is supplied regarding the unit using the test, the test ID number, the name of the part of the test (e.g., vocabulary) concerned, and the names of those responsible for the preparation of the testing and scoring materials. The bulk of the TIAF is for the entry of the following data:

a. the accession number
b. the item numbers of the base and scrambled (A & B) test versions
c. the key, the origin of each item---whether old or new, and, if old, whether an equating or non-equating item
d. the tests in which the items had previously been used
e. their item numbers in those old tests
f. a record of the items' relative difficulty on the new test as determined by the post-administration item and test analysis

g. the effectiveness of each item in ranking the test subjects on the part of the test (e.g., grammar) in which they are found, relative to the ranking effectiveness of that part of the test taken as a whole (r - bis.).

3. At the end of the TIAF are totals for each letter (A, B, C, D) keyed, to assure a sufficient distribution of each. The relative difficulty of that portion of the test is also indicated there, as is the correlation coefficient for that part as a whole.

4. Except for the last two columns, the TIAF should be filled in at the time of test construction. Following test administration and then item and test analysis of the results, the last two columns can be filled in. One such form---ultimately with all item accession numbers (see table by unit, in Memo-21) included---should be in the T&M Office files for each norming and placement examination for every unit.

Tests & Measurement Office

Hawaii University Language Center

Test Item-Assembly Form :

Sample Form

M-16

Unit _____

Test Form _____

Part of Test (if scored separately) _____

Name of Part _____

Test Assembled by _____

Key Prepared by _____

Key Reviewed by _____

ACCESSION NUMBER	BASE FORM NO.	SCRAMBLED FORM NO.	KEY	ORIGIN OF ITEM			PRIOR USE		△	r-bis
				OLD ITEM		NEW ITEM	TEST FORM	ITEM NO.		
				E	H-E					
	1									
	2									
	3									
	4									
	5									
	6									
	7									
	8									
	9									
	10									
	11									
	12									
	13									
	14									
	15									
	16									
	17									
	18									
	19									
	20									
	21									
	22									
	23									

P ____ of ____

(revised: October 1979)

ACCESSION NUMBER	BASE FORM NO.	SCRAMBLED FORM NO.	KEY	ORIGIN OF ITEM			PRIOR USE		△	r-bis
				OLD ITEM		NEW ITEM	TEST FORM	ITEM NO.		
				E	N-E					
	24									
	25									
	26									
	27									
	28									
	29									
	30									
	31									
	32									
	33									
	34									
	35									
	36									
	37									
	38									
	39									
	40									
	41									
	42									
	43									
	44									
	45									
	46									
	47									
	48									
	49									
	50									

No. of Items Keyed A ____
B ____
C ____
D ____

$M_{\triangle}$ = ____

$M_{r\text{-bis}}$ = ____

M - 17

From: Donald J. Malcolm

To: Unit Test Representatives

Tests and Measurement Office
Language Center
Kuwait University

21st February, 1979

Subject: Reading & Interpreting the Computer Output for LC Item Analyses

Below is an example of the computer output for one test item.

Item No.	A	B	C	D.	OMIT	NR	PT
1	153*	7	16	13	11	0	1.000
	76.500	3.500	8.000	6.500	5.500	0.000	
	52.051	38.002	43.583	43.831	45.718	0.000	

PPLUS	DELTA	R
0.765	10.111	0.511

The column headings designate the following:

1. A, B, C and D identify the four options in a four-choice question.

2. Omit refers to those individuals who did not respond to the question but who went on to answer subsequent questions.

3. NR is an abbreviation for "not reached". An individual is regarded as having not reached a particular question if he does not answer that question or any subsequent questions.

4. PT is an abbreviation for "proportion of all students who attempted the question". This proportion is equal to $\frac{N - NR}{N}$, where N is the total number of students tested.

5. P Plus (p+) refers to the proportion of individuals who attempted the question and who answered it correctly. Note that P+ is the same as the proportion of all individuals tested who obtained the correct answer only in the case of items that have been attempted by everyone. P+ is equal to $\frac{N(Correct)}{N-NR}$.

   P+ is used as an index of item difficulty. The more difficult the item the smaller the proportion of individuals who answer it correctly. However, there are two major disadvantages in using P+ in test development work. First, the scale runs the wrong way. Secondly, the scale is not an equal interval scale, i.e., a difference in P+ of .05 near the ends of the scale represents a greater difference in item difficulty than a difference of .05 near the middle of the scale. Hence, values of P+ cannot be legitimately averaged. Because of these limitations, delta (Δ) is used as the index of item difficulty.

BEST COPY AVAILABLE

6. Delta ranges in value from about 5.0 for very easy items to about 21.0 for very difficult ones. Delta values are related to P+ values through the normal curve. Values of P+ corresponding to values of delta are given in Appendix I.

7. R designates r-biserial, the coefficient of correlation between answering a question correctly and total score on the part of the test in which the question appears. It is a measure of the ability of an item to distinguish (or discriminate) between the better students (those who do well on the test as a whole) and the poorer students. Additional information pertaining to r-bis is given in Appendix I.

Under the columns headed A, B, C, D, Omit and NR, three rows of figures are given. These figures convey the following information:

Row 1. The numbers of students (frequencies) who chose responses A, B, C, or D or who did not attempt or reach the item. A+B+C+D+OMIT+NR = total number of individuals tested. The correct response to the item is designated by an asterisk (*).

Row 2. The percentages of all the students tested who chose responses A, B, C, or D, or who omitted or who did not reach the item. These percentages correspond to the frequencies given in row 1.

Row 3. The mean scores on the part of the test in which the item is included of those students who chose responses A, B, C or D, or who omitted or who did not reach the item. In arriving at this mean score, each student's raw score on the part of the test (the number of questions answered correctly) is converted to a standard score in a score distribution with a mean of 50 and a standard deviation of 10. The resulting standard scores will range range from a low of about 30 to a high of about 70. A mean score of about 50 indicates that the group of students is, as a whole, above average in ability.

Example of Interpretation of Item Analysis Data

The data given on page 1 may be interpreted as follows:

Some 153 of the students tested chose option A, the correct answer to the item (indicated by the asterisk after 153). Some 7, 16, and 13 students, respectively, chose distractors B, C, or D. A total of 11 students skipped the item. All 200 students, however, went on to answer subsequent questions. Hence, no individuals can be considered not to have reached the item (NR=0) and all 200 students can be regarded as having attempted the item (P+ = 1.00). The mean part score for the students who answered correctly was 52.051. Since this mean is above 50.000, it can be concluded that the students who chose the correct response were on the average superior (in terms of abilities measured by this part of the test) to those who chose the other options or who omitted the item. The mean scores of students who answered incorrectly

or who omitted the item are all below 50.000, indicating that these students were lower in ability than those who answered correctly.

Slightly more than three-quarters of the 200 students tested answered this question correctly (P+ = 0.765). This value of P+ corresponds to a delta of 10.111. The biserial-r for this item, 0.511, indicates a satisfactorily high correlation between answering these questions correctly and overall performance on the part of the test in which the question appeared.

The item analysis data for this item indicates that it is a relatively easy item that is functioning well. The better students tend to answer it correctly and the poorer students are lured away by the distractors. All distractors are functioning and none prove more attractive to better than to poorer students.

Appendix

Item Deltas and Biserial-r's

The most straightforward measure of the difficulty of an item is simply the proportion of students who attempt the item and get the right answer (p+). These proportions are reported in the analyses that are run on the Language Center tests. However, there are two undesirable features of p+ as a measure of item difficulty. One is that the scale of proportions runs in the opposite direction to that of difficulty (the easy items have larger p+ values than the harder ones). A more serious weakness of p+ is that the scale is not an equal interval scale. A difference between two p+ values of .10 near the center of the scale (say between .40 and .50) represents a smaller difference in the actual difficulties of the items than does the same differance at the ends of the distribution (say between .80 and .90). Because the p+ values do not constitute an equal-interval scale, mean p+'s and standard deviations of p+'s for a set of items cannot logically be calculated. To avoid these two limitations of p+, the delta scale ($\triangle$) of item difficulty is often used for test development work. Delta is related to p+ through the normal curve. The table below gives the values of p+ that correspond to values of delta.

Relationship between values of $\triangle$ and p+

		p+
very easy items	5	.98
	6	.96
	7	.93
easy items	8	.89
	9	.84
	10	.77
middle difficulty items	11	.69
	12	.60
	13	.50
	14	.40
	15	.31
hard items	16	.23
	17	.16
	18	.11

very hard items	19	.07
	20	.04
	21	.02

A second important statistic that is obtained for every item in an item analysis is r-bis, which is short for biserial r---a variety of correlation coefficient. Its possible values run from -1.00 through 0.00 to +1.00. It is a measure of the relationship between the scores that individuals receive on the single item (either a "1" for correct or a "0" for incorrect) and the total scores that the same individuals receive on the part of the test in which the item appears. If, say, 20 per cent of the students get the item correct and if these same students have the top 20 per cent of the scores on the part of the test in which the item is included, the relationship between answering the item correctly and performance on the part of the test as a whole is very high. Such an item is said to "discriminate" very well between the students with the higher part scores and those with the lower scores. In such a case as this, the r-bis will be very close to its maximum value of 1.00 (a value never attained in practice). If the students who "pass" the item do no better on the test as a whole than those students who miss the item, then the item has failed to separate the better students from the poorer. In this case the discriminating power is zero and the r-bis is zero. Occasionally, an item will have a negative r - bis. This implies that the poorer students rather than the better ones tend to choose the "correct" response. Almost invariably this situation indicates a defective item. Either the item has been incorrectly keyed (i.e. the wrong option is specified as the right answer) or one or more of the distractors is for some reason leading the better students away from the intended answer.

An achievement test or a placement test is by definition a device for making discriminations among individuals with respect to what they have learned. Accordingly, if the test is to be an efficient one, each item in it should be capable of distinguishing to some degree the better students from the poorer ones. Unfortunately---from the point of view of ease of interpretation of the results of item analyses---there are no fixed rules for deciding whether the biserial-r for a particular item is sufficiently high to justify its continued use in a test or sufficiently low to justify its rejection. If the items in a test are very homogeneous with respect to the abilities they measure, the biserials will run high---say, in the .60s and even up in the .70s. In a vocabulary test, for example, that is a sample of high-frequency words introduced in a specific course, the biserial-r will be high. In a vocabulary test made up of low-frequency words selected from different subject areas such as biology and literature, the biserial will be much

lower because of the heterogenity of the content represented by the items. Reading comprehension items usually have lower biserials than structure items because of the wider variety of skills measured by the reading comprehension items than by the structure items.

In general, items with biserials below .30 should be examined for possible flaws. It may be that one or two distractors are not functioning and revision is needed. The item analysis data will be most helpful in this inspection. Items that have low biserials and high deltas are usually faulty. The high delta comes about because the item may have a double key (two correct answers) or it may have no acceptable answer. Consequently few students choose the intended response (hence, the high delta indicating a difficult item); and there may be no relationship between answering this question "correctly" and performance on the rest of the test (hence the low biserial).

Occasionally an item will have a very low biserial-r, but inspection reveals no apparent flaws in it. It may be that the item measures something different than what is measured by the other items in the test of which it is a part. Consequently student performance on that item may not correlate highly with the performance on the remaining items. Such an item---if the skill it measures is important---may not contribute to the internal consistancy reliability of the test but it will contribute to the validity of the test as a predictor of language performance. For this reason, one should guard against throwing out items without looking at them simply because their biserials fall below some arbitrary value.

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-18 | مكتب الاختبارات والقياس

Date: 28th Oct., 1979 التاريخ: _____

Ref : الرقم: _____

MEMO

From Donald J. Malcolm

To Unit Test Representatives

Subject Instructions for Completing the "Summary Form - Item and Test Analysis Data"

A copy of the Summary Form is attached to this memorandum. All the information required to complete this form is obtainable from the content specifications for the test and from the computer output for the item and test analyses. In making out the form, please note the following:

1. Heading (page 1)

   A. Form Designation (Self-explanatory)

   B. Test Population

      Identify the group of students who took the test. For example, students entering the College of Arts, or students completing the 098 course in the College of Engineering.

   C. Language Unit and Administration Date (Self-explanatory).

2. Test Composition and Timing (page 1)

   A. Part, Name of Part, Number of Items, Item Type, and Time allowed can be obtained from the test book or the test specifications. In specifying "Item Types" use the codes given in the T&M Memo on "Codes for Item Types".

   B. The "notes" section should be used to describe any unusual scoring procedures for multiple-choice items (such as the differential weighting of

responses or the application of penalties for guessing). Also to be noted here are differential weightings of part scores, if any, in determining the total score.

3. Obtained Test Statistics (page 1)

A. Mean and S.D. (Standard Deviation).

These statistics for the part and total scores are given in the computer output directly after their respective score distributions.

B. Median

The median (which is the point above which and below which fifty percent of the scores in a distribution lie) is not given in the LC computer output. See any elementary statistic book on how to calculate the median or ask someone in the T&M Office to show you how to do it.

C. Range

The range of scores for a distribution is from the lowest score obtained by anyone to the highest. If the lowest score anyone received on a given part of a test is 6 and the highest is 35, the range is 6 - 35. This statistic is obtained by inspecting the score distribution reported in the computer output.

D. r_{xx}

The coefficient of reliability for each part of the test and for the total is given in the computer output directly after the corresponding score distribution. It is an internal consistency coefficient calculated on the basis of Kuder - Richardson formula 20.

E. SE_m

This statistic is the standard error of measurement. It is not reported in the L.C. computer output. However, it can be quickly calculated from

$$SE_m = SD\sqrt{1 - r_{xx}}.$$

For test scores near the average score, it gives an indication of how much confidence we can have in a particular score. For scores in the middle range, the chances are two out of three that the true scores lie between the obtained score plus or minus one SE_m. The chances are about 95 out of 100 that the true score lies between that obtained plus or minus two SE_ms .

4. Expected Means for Tests of Middle Difficulty and Chance Score Data (page 1)

A. M_{Exp}

If a multiple-choice test is constructed to be of middle difficulty for the group of students that will take it, the expected mean score (M_{exp}) can be determined from the following relationship

$$M_{exp} = .50n + .50M/s = .50n\ (1 + 1/s)$$

where n = the number of items and s = the number of options. For a test made up of 40 four-choice items, $M_{exp} = .50 \times 40\ (1 + 1/4) = 25$.

This relationship is based on the assumption that if the test is of middle difficulty, students will answer one half of the questions correctly because they know the answers. In guessing on the remaining half, they will on the average guess 1/s of them correctly.

B. M_c

M_c designates the mean chance score. It is the average score that would be obtained on a multiple-choice test by a group of individuals who know absolutely nothing about the subject matter and who guess the answer to every question. It is equal to 1/s times the number of items, where s = the number of options. In a test of 40 four-choice questions, $M_c = 1/4 \times 40 = 10$.

C. SD_c

SD_c designates the standard deviation of the distribution of chance scores.

It is equal to $\sqrt{n\ p\ q}$,

where n = the number of questions, p = the probability of getting one item correct by guessing, and q = the probability of getting the item wrong (q = 1 - p).

In a test of 40 four-choice items, $SD_c = \sqrt{40 \times .25 \times .75} = 2.74$.

Accordingly, about two-thirds of all chance scores will fall within the range 7 to 13 ($M \pm 1\ SD = 10.0 \pm 2.74$).

D. $M_c + 1.65\ SD_c$

This gives the score below which 95% of scores obtained purely by chance will lie. For the 40-item, four-choice test, this score is 10 + (1.65 X 2.74) = 14.52, or 15, to round off to the nearest possible score. The fact that an obtained score is below 15 does not necessarily mean that that score is a chance one. It might be valid, but one cannot be reasonably certain. The possibility that any score above 15 is a chance score is most unlikely.

E. % N in Chance Range

This is simply the percentage of students whose scores fall within the chance range defined in D above. It is obtained from the frequency distributions of part and total scores in the computer printout.

5. Item Statistics: Frequency Distributions of Deltas and r-biserials (page 2)

A. Deltas

(1) Using the item analysis data given in the computer output, obtain the frequency distribution of deltas for each part of the test and enter these figures in the columns provided.

(2) For each range of deltas specified in the left column, add the numbers of items in that range in each part and enter the total in the right-hand column.

(3) The means and standard deviations of the delta distributions for the parts of the test are given in the computer output. However, these statistics are not given for the test as a whole. The mean delta for all items in the test is the weighted mean of the means of the parts. This is obtained by multiplying for each part the mean delta for that part by the number of items in it, summing these products over the several parts, and dividing the resulting sum by the number of items.

(4) The standard deviation of the deltas for the test as a whole will be calculated by the T&M Office.

B. r-biserials

(The same procedures described above for deltas are to be followed in connection with the r-biserials.)

6. Speededness Data (page 3)

Note: If scrambled versions of the same test form are given at a particular administration, the speededness data called for should be for the base form only (usually designated "A," "1," or "A1"). The data should not be obtained from the computer output for the combined scrambles. The reason for this is that the positions of specific items in the scrambles vary. Hence the computer output for the combined scrambles gives an erroneous picture of the speededness characteristics of the test.

A. Number of Items Finished by All Students

For each part of the test, this number is ascertained by inspecting the column labeled "PT" (for the proportion trying) in the item analysis for that part. The number of items finished by all students is the last item for which PT = 1.00.

B. Percentage of Students Who Finished All Items

The percentage of students finishing, or completing, all the items in a part is 100 times the PT value given for the last item in a part. If the PT value for the last item is .927, the percentage of students completing that part is 100 X .927 = 92.7%.

C. Number of Items Finished by 80% of the Students

This number is the number of the last item in a part that has a PT value greater than 0.799.

7. Part-Score Intercorrelations (page 3)

A. Obtained

The obtained intercorrelations among the part scores and between the part scores and total scores are given at the end of the computer output for the test analysis.

B. Corrected for Attenuation

A coefficient of correlation corrected for attenuation is an estimate of what the correlation between two sets of scores would be if each set were perfectly reliable. It is obtained by dividing the obtained r by the square root of the product of the reliabilities of the two sets of scores.

$$r_{1T\,2T} = \frac{r_{1\,2}}{\sqrt{r_{1\,1}\quad r_{2\,2}}}$$

In effect, the corrected r's adjust for differences in the reliabilities of the different parts as a consequence of the different numbers of items they contain.

8. Test Analysis Recommendations

This part of the report will be completed by the T&M Office and discussed with the UTR.

T & M Office
Language Centre
Kuwait University

SUMMARY FORM: ITEM AND TEST ANALYSIS DATA

Form Designation _____ Faculty _____
Test Population _____ Administration Date _____

Test Composition and Timing

Part	Name of Part	Number of Items	Item Type	Time Allowed
___	_____	_____	_____	_____
___	_____	_____	_____	_____
___	_____	_____	_____	_____
___	_____	_____	_____	_____
___	_____	_____	_____	_____
TOTAL		_____		_____

Notes _____

Obtained Test Statistics

Part	N	Mean	S.D.	Median	Range	r_{xx}	SEm
___	___	___	___	___	___	___	___
___	___	___	___	___	___	___	___
___	___	___	___	___	___	___	___
___	___	___	___	___	___	___	___
___	___	___	___	___	___	___	___
TOTAL	___	___	___	___	___	___	___

Notes _____

Expected Means for Tests of Middle Difficulty and Chance Score Data

Part	$M_{EXP.}$	M_C	SD_C	$M_C+1.65SD_C$	% N in Chance Range
___	___	___	___	_____	_____
___	___	___	___	_____	_____
___	___	___	___	_____	_____
___	___	___	___	_____	_____
___	___	___	___	_____	_____
TOTAL	___	___	___	_____	_____

Item Statistics - Frequency Distributions

Deltas

Obtained Deltas	Part I	Part II	Part III	Part IV	Part V	Total
19+						
18-18.99						
17-17.99						
16-16.99						
15-15.99						
14-14.99						
13-13.99						
12-12.99						
11-11.99						
10-10.99						
9-9.99						
8.8.99						
7-7.99						
6-6.99						
5-5.99						
No. of items						
Mean Delta						
SD of Delta						

r - biserial

Obtained r - bis	Part I	Part II	Part III	Part IV	Part V	Total
.90-.99						
.80-.89						
.70-.79						
.60-.69						
.50-.59						
.40-.49						
.30-.39						
.20-.29						
.10-.19						
.00-.09						
No. of Items						
Mean r-bis						
SD of r-bis						

Speededness Data

	Part I	Part II	Part III	Part IV	Part V	
1.	___ /	___ /	___ /	___ /	___	number of items finished by all students
2.	___ /	___ /	___ /	___ /	___	per cent of students finished all items
3.	___ /	___ /	___ /	___ /	___	number of items finished by 80% of the students

Part Score Intercorrelations

Obtained

Part	1	2	3	4	5	
1	X					
2	___	X				
3	___	___	X			
4	___	___	___	X		
5	___	___	___	___	X	
Total	___	___	___	___	___	X

Corrected for Attenuation

Part	1	2	3	4	5	
1	X					
2	___	X				
3	___	___	X			
4	___	___	___	X		
5	___	___	___	___	X	
Total	___	___	___	___	___	X

Test Analysis Recommendations

Signed _____

Faculty _____

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-20 | مكتب الاختبارات والقياس

Date: 17th October, 1979 — التاريخ :

Ref : — الرقم :

MEMO

From Donald J. Malcolm

To UTRs

Subject Item Cards for the T&M File
(format revised: October 1979)

1. The item card (see sample, attached) records essential information about the history of usage and extent of relative effectiveness of a particular multiple-choice test item. Once an item has been altered in any way, however little, it should be considered an entirely new item and placed on a new card.

2. The item card carries the following information about an item on its obverse side:

   a. the part of a test in which it was used (e.g., vocabulary)

   b. the language unit(s) in whose exams it was used (e.g., English: Arts)

   c. the lowest course - level for which it can be considered an appropriate and representative test item for purposes of course-final (achievement) exams

   d. the item accession number, a hyphenated, seven-digit identification number which is unique for every item in the T&M Office files

   e. the history of usage of the identical item on all previous LC - level standardized exams for which statistical analysis was subsequently undertaken

   f. the item itself as reproduced (or cut out) from the test form

3. The reverse side of the item card carries the following detailed information for each administration of an item:

a. its history of prior usage on various *LC*-level test forms

b. the number of subjects to whom that exam was administered (N)

c. the number of subjects who selected each option on the item

d. the correct option (starred)

e. the number of subjects who reached but did not attempt the item (omit)

f. the number of subjects who did not reach the item (N.R.)

g. the percentage of subjects who attempted the item (P total)

h. the number of the item on this particular test form

i. the means on the part-scores, for this part of the test, of subjects who selected each of the four options (M.A, M.B, etc.)

j. the means on the part-scores, for this part of the test, of subjects who either omitted the item (M.O) or did not reach it (M. NR)

k. the percentage of subjects who attempted the item and responded correctly (P+)

l. the degree of difficulty of the item relative to other items in the same part of the test (△)

m. the performance of the item, in ranking subjects for ability, when compared with the same performance of that portion of the test (e.g., grammar) as a whole, of which the item is a part (r - bis).

4. All items deemed acceptable on the basis of item and test analyses are stored in the T&M Office item bank, available for instant retrieval and insertion into *LC*-level exams. Once an item has been so determined to be effective in discriminating among varying levels of ability in a particular type of test, it becomes a permanent part of the bank, available for the construction of even more sensitive testing instruments. It is then available for use either in aiding in the diagnosis of ability-levels of incoming students for placement purposes, or in measuring student achievement in language-unit course-final exams.

5. Use of the hyphenated, seven-digit accession numbers not only assigns a unique identification code to each item in the file. It also permits serial ordering of the items for purposes of ready retrieval, and thus rapid assembly of new test forms. This assembly is facilitated via the accession numbers, through cross reference with the Test Item-Assembly Forms, which constitute a detailed record of each administration of a given test form.

6. The only cards in the files not to be so numbered serially (because they do not contain items) or to be otherwise filled out with test statistics are those providing essential supplementary information basic to understanding the administration of the items on the other cards. Examples of such auxiliary "non-statistical" cards include 1) those containing reading comprehension passages, and 2) those atypical sections of the exams of some units, tested with a multiple-choice format, such as "paragraph structure" (sentence ordering, in which students are presented the scrambled sentences of a paragraph and must re-order them logically).

Gr LC PS RC Sp Vo Wr _____

_____

_____

ALU _____

ELU: A C E L S

FLU

COURSE _____

(lowest for which this item is appropriate)

PRIOR USE

FRONT

FORM	BASE N	A	B	C	D	OMIT	N.R	P TOTAL	△SCALE	△ E	CRITERION
	ITEM NO	M.A	M.B	M.C	M.D	M.O	M.NR	M TOTAL	P +	△O	R. BIS.

FORM	BASE N	A	B	C	D	OMIT	N.R	P TOTAL	△SCALE	△ E	CRITERION
	ITEM NO	M.A	M.B	M.C	M.D	M.O	M.NR	M TOTAL	P +	△O	R. BIS.

FORM	BASE N	A	B	C	D	OMIT	N.R	P TOTAL	△SCALE	△ E	CRITERION
	ITEM NO	M.A	M.B	M.C	M.D	M.O	M.NR	M TOTAL	P +	△O	R. BIS.

FORM	BASE N	A	B	C	D	OMIT	N.R	P TOTAL	△SCALE	△ E	CRITERION
	ITEM NO	M.A	M.B	M.C	M.D	M.O	M.NR	M TOTAL	P +	△O	R. BIS.

BACK

-ST/

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-21 | مكتب الاختبارات والقياس

Date: October 20, 1979 التاريخ :

Ref : الرقم :

MEMO

From Victor W. Mason

To Unit Test Representatives

Subject Item Accession-Number Table

1. The attached table explains the item-accession numbering system that is to be used in connection with the T&M Office item files. The appropriate numbers are to be entered on all item cards and on the Test Item-Assembly Form that is prepared for each new test (see M-16).

2. Each item's accession number is divided into a two-digit language and unit pre-code and a five-digit post-code serially numbering each item within a unit up to a possible 99,000 items and up to 5000 items for each category. Thus, in the following example accession number:

22 - 15363

the first digit, 2, denotes the English language program (1 is for Arabic and 3, French). The second digit, 2, indicates Commerce, the second ELU in alphabetical order (1 is for Arts; 3, for Engineering; 4, for Law, and 5, for Science).

3. The entire coding system of accession numbers can be seen in Memo 21. In the case of each unit, provision for possible new types of items in the future is made under the category entitled "(other)".

EE (English: Engineering): 23-00001 --- 23-40000

Gr:	23-00001 --- 23-05000
LC:	23-05001 --- 23-10000
PS:	23-10001 --- 23-15000
RC:	23-15001 --- 23-20000
Sp:	23-20001 --- 23-25000
Vo:	23-25001 --- 23-30000
Wr:	23-30001 --- 23-35000
(other):	23-35001 --- 23-40000

EL (English: Law): 24-00001 --- 24-35000

Gr:	24-00001 --- 24-05000
LC:	24-05001 --- 24-10000
RC:	24-10001 --- 24-15000
Sp:	24-15001 --- 24-20000
Vo:	24-20001 --- 24-25000
Wr:	24-25001 --- 24-30000
(other):	24-30001 --- 24-35000

ES (English: Science): 25-00001 --- 25-35000

Gr:	25-00001 --- 25-05000
LC:	25-05001 --- 25-10000
RC:	25-10001 --- 25-15000
Sp:	25-15001 --- 25-20000
Vo:	25-20001 --- 25-25000
Wr:	25-25001 --- 25-30000
(other):	25-30001 --- 25-35000

F (French): 30-00001 --- 30-35000

Gr:	30-00001 --- 30-05000
LC:	30-05001 --- 30-10000
RC:	30-10001 --- 30-15000
Sp:	30-15001 --- 30-20000
Vo:	30-20001 --- 30-25000
Wr:	30-25001 --- 30-30000
(other):	30-30001 --- 30-35000

Item Accession-Number Table

1. The first digit of each 7-digit number indicates the language:

   1: Arabic 2: English 3: French

2. The second digit indicates the unit concerned:

   0: No differentiation
   1: Arts
   2: Commerce
   3: Engineering
   4: Law
   5: Science

3. The third through seventh digits are for serially ordering the items in each program, up to a maximum of 5000 items per item-type (e.g., grammar, vocabulary, etc.).

****

A (Arabic): 10-00001 --- 10-35000

Gr:	10-00001 --- 10-05000
LC:	10-05001 --- 10-10000
RC:	10-10001 --- 10-15000
Sp:	10-15001 --- 10-20000
Vo:	10-20001 --- 10-25000
Wr:	10-25001 --- 10-30000
(other)	10-30001 --- 10-35000

EA (English: Arts): 21-00001 --- 21-35000

Gr:	21-00001 --- 21-05000
LC:	21-05001 --- 21-10000
RC:	21-10001 --- 21-15000
Sp:	21-15001 --- 21-20000
Vo:	21-20001 --- 21-25000
Wr:	21-25001 --- 21-30000
(other)	21-30001 --- 21-35000

EC (English: Commerce) 22-00001 --- 22-35000

Gr:	22-00001 --- 22-05000
LC:	22-05001 --- 22-10000
RC:	22-10001 --- 22-15000
Sp:	22-15001 --- 22-20000
Vo:	22-20001 --- 22-25000
Wr:	22-25001 --- 22-30000
(other)	22-30001 --- 22-35000

Key: Gr: Grammar Vo: Vocabulary
LC: Listening Comprehension Wr: Writing
RC: Reading Comprehension (other): other kinds of items
Sp: Speaking

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-22 | مكتب الاختبارات والقياس

Date : October 1979 | التاريخ :

Ref : DJM/st | الرقم :

MEMO

From Donald J. Malcolm

To Unit Test Representatives

Subject Code Designations for Various Item Types

The item codes given below are to be used in the test content specifications and in other T&M documents in which reference is made to the kinds of items included in particular tests. In using these codes, please note the distinction that is made in the item-writing memo (No. 12) on "Items Primarily Intended to Measure Listening or Reading Comprehension Skills and then Intended to Measure a Complex of 'Integrated' Skills".

The listing that follows is not exhaustive but covers the item types most frequently used in LC final and placement tests.

Terms used in describing the various item types are defined as follows:

Stimulus Material:

The printed or spoken words that are presented to the student and to which he is required to react in some prescribed way.

Cognitive Task:

The mental operations or processes that the student is expected to perform on the stimulus material.

Response Modality:

A class of overt behaviours by means of which the student conveys to the examiner his response to the stimulus material. Response modalities include but are not limited to (1) non-verbal acts, (2) oral responses, (3) written responses ("supply") and (4) selection among options presented ("multiple-choice").

In designating the item codes certain principles have been followed. These are

1. The first letter or two designates the major skill area (for example. RC = reading comprehension, V = vocabulary).

Date :

التاريخ :

Ref :

الرقم :

MEMO

From

To

Subject

2. The last letter or the last combination of letters indicates the response modality (S = supply; 4MC = multiple choice with 4 options; O = response other than supply or multiple choice).

3. The middle letter or letters identify the type of stimulus material (for example, Q = discrete questions, G = graphical material).

Example:

G SCP S

- G — a grammatical structure item
- SCP — sentence completion with prompts
- S — supply-type responses

1. Items Intended Primarily to Measure Listening Comprehension

	Stimulus Material	Cognitive Task	Response Modality	Item Code
A.	Discrete sentences	-paraphrase -logical inference	-multiple choice -supply -other	LCS-4MC LCSS LCSO
B.	Discrete questions	-logical answer	-multiple choice -supply -other	LCQ-4MC LCQS LCQO
C.	Dialogues	-answer q's	-multiple choice -supply -other	LCD-4MC LCDS LCDO
D.	Passages	-answer q's	-multiple choice -supply -other	LCP-4MC LCPS LCPO

2. Items intended primarily to measure reading comprehension

A.) as above for listening comprehension
B.) but with RC substituted for LC
C.) in the item code
D.)

E.	Graphs	-answer q's	-multiple choice -supply -other	RCG-4MC RCGS RCGO
F.	Tables	-answer q's	-multiple choice -supply -other	RCT-4MC RCTS RCTO

3. Items included primarily to measure writing ability directly
(all items in this category require written responses)

	Stimulus Material	Cognitive Task	Item Code
A.	Discrete questions	formulate logical replies in complete sentences using prompts	WSSP
B.	Discrete questions	formulate logical replies in complete sentences without prompts	WSSNP
C.	Discrete sentences	formulate questions, using prompts, for which the SM constitutes the appropriate answers	WQP
D.	Discrete sentences	(as in C but without prompts)	WQNP

Stimulus Material	Cognitive Task	Response Modality	Item Code
E. Guide questions or prompts	write paragraphs		WPP
F. Assigned topic, no prompts	write paragraphs		WPNP
G. Assigned topic	write series of paragraphs		WENP

4. Items intended to measure vocabulary

Stimulus Material	Cognitive Task	Response Modality	Item Code
A. Discrete sentences with gaps	insert word or words to complete	-multiple choice -supply -other	VC-4MC VCS VCO
B. Discrete sentences with words underlined	give synonyms for underlined words	-multiple choice -supply -other	VS-4MC VSS VSO
C. Discrete questions calling for examples related to word being tested	give examples	-multiple choice -supply -other	VE-4MC VES VEO
D. Vocabulary in context of paragraphs	identify lexical item defined or referred to	-multiple choice -supply -other	VP-4MC VPS VPO

5. Items intended to measure control of grammatical structures

Stimulus Material	Cognitive Task	Response Modality	Item Code
A. Discrete sentences with gaps and prompts	insert word or words to complete	-multiple choice -supply -other	GSCP-4MC GSCPS GCPO
B. Discrete sentences with gaps but without prompts	insert word or words to complete	-multiple choice -supply -other	GSC-4MC GSCS GSCO
C. Sentences with gaps in paragraphs, prompts provided	insert word or words to complete	-multiple choice -supply -other	GPP-4MC GPPS GPPO

Stimulus Material	Cognitive Task	Response Modality	Item Code
D. Sentences with gaps in paragraphs, no prompts	insert word or words to complete	-multiple choice -supply -other	GP-4MC GPS GPO
E. Two-sentence dialogue with gap in second person's contribution	treat as in A or B above		
F. Discrete sentences or questions with final words or phrases to be supplied by student (sentence completion)	treat as in A or B above		
G. Discrete sentences or questions with several words or phrases underlined, one of which may be grammatically incorrect (error recognition)	identify ungrammatical portion of sentences, if any	-multiple choice -supply -other	GER-4MC GERS GERO
H. As in G above but without underlining	identify ungrammatical portion, if any	-multiple choice -supply -other	(N.A.)* GERS GERO
I. Discrete sentences given with transformational instructions (supply-type) or without such instructions (multiple choice)	perform transformation (supply-type) or recognize correct transformation (multiple-choice)	-multiple choice -supply -other	GT-4MS GTS GTO

6. Items intended to measure "integrated" skills

A. Items in which listening and writing skills are integrated:

Stimulus Material	Cognitive Task	Response Modality	Item Code
1. Dictated material	reproduce dictated material	-multiple choice -supply -other	(N.A)* IDS (N.A)*

* not applicable

Stimulus Material	Cognitive Task	Response Modality	Item Code
2. Material presented orally	produce summary	-multiple choice -supply -other	(N.A.)* IOWSS (N.A.)*

B. Items in which reading comprehension and writing skills are integrated:

Stimulus Material	Cognitive Task	Response Modality	Item Code
1. Reading passages followed by open-ended questions requiring answers in one or more complete sentences	write answers	-multiple choice -supply -other	(N.A.)* IRCWS (N.A.)*
2. "True" cloze passages with every nth word deleted. (pseudo-cloze passages in which particular classes of words have been omitted are to be classified as vocabulary or structure items in context)	insert missing words	-multiple choice -supply	IC-4MC ICS

* not applicable

-ST/

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-23 | مكتب الاختبارات والقياس

Date : February 26, 1980 — التاريخ : .....

Ref : DJM/st — الرقم : .....

MEMO

From Donald J. Malcolm

To Director, Assistant Director, Unit Supervisors, Course Coordinators & UTRs

Subject Addition of M-23 to Test Development Manual

1. Please add this newest memo, entitled "Test Cover Sheet," to your Test Development Manual. It is designated M-23.

2. The enclosed form is a sample of a suitable format for the cover sheet that unit supervisors are to use in submitting final-test copy for final review by the T&M Office. This format has already incorporated recent suggestions that TD committees be expanded to include UTRs and that reviewers include instructors both teaching and not teaching the course for which a given final has been prepared.

3. An essential part of the review process is to have the reviewers first take the test as a student would (see M-3, p.4: No.12). Critical comments by reviewers are also to be made in conjunction with this "pre-test" and the TD committee is then to take into account the results of both the pre-test and the reviewers' incidental critical comments.

COLLEGE OF ENGINEERING
ENGLISH LANGUAGE UNIT
TEST COVER SHEET

_____
(Date)

1. **Course Identification** _____

2. **Estimated Number of Students Who Will Take The Test** _____

3. **Number of VERSIONS** _____

4. Test Development Committee Members

   a) _____ , Chairperson
   b) _____
   c) _____
   d) _____ , (UTR)
   e) _____
   f) _____

5. Test Reviewers

   a. Teachers of the Course

      1) _____
      2) _____

   b. Other Teachers

      1) _____
      2) _____

In my opinion, this test constitutes an appropriate final test for the course for which it has been developed.

_____
SUPERVISOR

Memo for: ELU Supervisors, UTR's, Course Coordinators and Teachers: M-24

From: Donald J. Malcolm

Date: May 4, 1980

Subject: Procedures for Calculating and Using Standard Scores

1. Some reasons for using standard scores

Standard scores are particularly useful in situations in which it is necessary to combine several test scores in order to arrive at a single, composite score. Two such situations that occur within the ELU's are these:

(1) the combining of scores on several unit tests to determine an overall score for performance on such tests, and

(2) the combining of scores on several components (course-work, unit tests and final examination, for example) to determine final grades.

The basic problem associated with simply averaging the scores on, say, several unit tests to obtain an overall average is that this procedure disregards differences in the means and standard deviations of the score distributions for the different tests. As a result of this, the unit test with the highest mean score will have a greater effect than the others in determining the mean of the distribution of composite scores. Similarly, the unit test with the greatest variability in obtained scores (i.e., with the largest standard deviation) will have a greater effect than the others on the variability of the composite scores. The advantage of converting

scores on several unit tests, or on the components entering into the determination of final grades, into standard scores is that the means and standard deviations of the constituent score distributions are respectively equalized. Hence, all have an equal weight, or effect, in the determination of the composite scores. In the situation where a priori weights are applied (as in calculating final grades from component scores), the use of standard scores assures that the a priori weights are truly reflected in the resulting distribution of composite scores.

2. Standard scores appropriate for LC purposes

Any obtained score distribution can be converted into a corresponding standard score distribution with any mean and standard deviation that one cares to assume. For example, if a test is given that results in a score distribution with a mean of 90 and a standard deviation of 15, the test scores can readily be converted so as to yield a score distribution with a mean of 50 and a standard deviation of 10. The transformation from one scale to the other has no effect on the rank-ordering of individuals. The person who is fifth from the top on the original scale will be the person fifth from the top on the converted scale. For Language Center purposes, it is recommended that obtained score distributions (as for the unit tests) be converted to standard distributions with a mean of 50 and a standard deviation of 10 (as in the example cited above).

3. Using standard scores in connection with unit tests

A. Preliminary considerations

(1) When standard scores are used, the effort should not be made to assemble tests that yield scores on an absolute scale from, say, 60 per cent correct to 100 per cent correct. A difficulty with this approach is that only 40 per cent

of the total possible marks is used to make distinctions among all those (the large majority) who will receive passing grades. The remaining 60 per cent of the total marks simply discriminate among the failures and, hence, serve no useful purpose. In the case of tests that are to be used for assigning grades, it is better to have the tests pitched at middle difficulty for the students who will take them. A test made up of supply-type items (essay, fill-in-the gaps, etc.) is of middle difficulty when the average score is about 50 per cent of the total possible marks. A test made up of multiple-choice questions with four options is of middle difficulty when the average score is about 62 per cent of the total number of such items. Tests of middle difficulty tend to distribute students over the whole range of possible scores, and, by so doing, permit a more reliable assignment of final grades than is possible when the test scores are compressed into a relatively narrow range from 60 to 100 percentage points.

(2) In the case of Unit Tests that are to be given to several sections of the same course and that contain substantial amounts of response material that must be subjectively graded, panel marking is much to be preferred over the reading of papers for a particular section by the teachers of that section. The reason for this, of course, is to reduce the inevitable effect of differences in the frames of reference and standards of individual teachers. Standard scores based on the performance on a Unit Test of <u>all</u> students from all sections of a course will not compensate for differences in the grading practices of individual teachers when each teacher grades his or her students' papers.

4. Calculation of Standard Scores

The conversion of an obtained test score to a standard score in a distribution with a mean of 50 and a standard deviation of 10 involves only the multiplication of the obtained score by a constant A and adding a constant B to the resulting product, as is shown below:

$$X_{SS} = A X_{OB} + B$$

where X_{OB} is an obtained test score

X_{SS} is the corresponding standard score

$A = \frac{10}{SD_{OB}}$, where SD_{OB} is the standard deviation of the raw score distribution

$B = (50 - \frac{10}{SD_{OB}} M_{OB})$, where M_{OB} is the mean of the raw score distribution

Example: A unit test is administered to all the students in all the sections of a course. The mean and standard deviation of the resulting raw score distribution are then calculated. These statistics may be obtained directly on many "scientific" pocket calculators and on the HP 97 in the T&M Office. Assume that the mean of the distribution (M_{OB}) is 30.0 and that its standard deviation (SD_{OB}) is 8.2. Constants A and B are then determined as shown below:

$$A = \frac{10}{SD_{OB}} = \frac{10}{8.2} = 1.22$$

$$B = (50 - \frac{10}{SD_{OB}} M_{OB}) = (50 - 1.22 \times 30)$$
$$= (50 - 36.60)$$
$$= 13.40$$

Hence the conversion equation becomes

$$X_{SS} = 1.22 X_{OB} + 13.40$$

If, in the example given above, a student has an obtained score of 20, his standard score is obtained by multiplying that score by 1.22 and adding 13.40 to the product.

$$X_{SS} = 1.22 \times 20 + 13.40 = 22.40 + 13.40$$
$$= 35.80$$

In calculating standard scores on individual tests, the standard score should be rounded off to the nearest whole number, i.e., 35.8 would be reported as 36. In the case of means of several standard scores (as in determining composite final grades based on several components), the mean score should be reported to the first decimal place. Carrying out the scores to more decimal places implies a degree of precision that does not exist in the field of language testing.

By use of a pocket calculator that can be programmed, the conversion of obtained scores to standard scores can be done very quickly. The HP 97 in the T&M Office is available for use in the office for this purpose.

5. <u>The Range of Standard Scores</u>

If the distribution of obtained scores is normal, i.e., the distribution curve is bell-shaped and symmetrical about the mean---as is usual with tests that are approximately of middle difficulty ---the following percentages of standard scores will fall within the ranges indicated.

Table 1

Standard Score Range	Percentage of Scores in Range
80 and above	0.13 %
70 to 79	2.15 %
60 to 69	13.59 %
50 to 59	34.13 %
40 to 49	34.13 %
30 to 39	13.59 %
20 to 29	2.15 %
19 and below	0.13 %

It will be noted in the above table that nearly all standard scores will be in the range from 20 to 80. Unless the test on which the scores are based was extremely easy or extremely difficult, very few standard scores (about one in a thousand) will be above 80 or below 20.

In using standard scores care should be taken to label them as such so that they will not be confused with percentage scores. A score of 80 on the percentage scale is only mediocre. On the standard score scale, 80 is extremely high.

6. Averages (Means) and Weighted Averages of Standard Scores

Averages and weighted averages for several standard scores are obtained in the same manner as for any other set of scores. Given below are the standard scores received by a student on three unit tests, X, Y and Z.

	Standard Scores on Unit Tests		
	X	Y	Z
Student No. 5572	38	42	46

If each score is to receive equal weight in a composite total score, add the three scores and divide by three.

$$\frac{38 + 42 + 46}{3} = 42.0 \quad \text{(Simple mean)}$$

If for some reason, it is considered desirable to give the score on test Z twice the weight of either X or Y, multiply each score by its weight, sum these products, and divide by the sum of the weights, viz.,

$$\frac{(1) \times 38 + (1) \times 42 + (2) \times 46}{4} = 43 \quad \text{(weighted mean)}$$

7. Assigning Standard Scores for Classwork

In the Language Center final grades are generally based on three components---unit tests, classwork and final examinations. In the case of the unit tests and final examinations, standard scores can be calculated directly from the obtained score distributions. The application of standard scores to student performance on classwork, however, represents a much different situation in that there are no common measure of classroom performance and, consequently, a much higher degree of subjectivity enters into grades on this variable than on those based on tests. Perhaps the best thing that can be done under these circumstances is to ask individual techers to assign standard scores to students in their sections roughly according to the distribution given in Table 1. In the case of very small sections, say of fewer than 10 students, considerable variations from the suggested distribution are to be tolerated because of the operation of chance factors. In the case of two sections under the same teacher with say 50 students, the distribution of grades for classwork should approximate quite closely those given in Table 1.

8. Calculating the Final-grade Standard Score from the Standard Scores of the Components

The procedure for obtaining the weighted mean score of the final-grade components is the same as that described

in section 6 of the memorandum for obtaining the weighted mean of the standard scores for several unit tests.

The example that follows is based on the assumption that performance on the unit tests is to count 40 per cent of the final grade, classwork 20 per cent and the final examination 40 per cent.

	Standard Scores for		
	Unit Tests	Classwork	Finals
Student No. 08266	45	38	40

Inasmuch as relative weights of 40%: 20%: 40% are the same as 2 : 1 : 2, the weighted mean of the above test scores is given by

$$\text{SS (Total)} = \frac{(2) \times 45 + (1) \times 38 + (2) \times 40}{5} = 41.6$$

It is recommended that the above score be multiplied by 10 and reported as a three-digit number, 416. This will clearly identify the total score as such. It will prevent confusion with the absolute percentage scale from 0 to 100, and it will provide for finer distinctions among large numbers of students (even through the distinction based on the third digit may be more a function of chance than of reliable measurement).

9. Assigning Letter Grades Based on the Final Grade Standard Scores

The principle is simple enough. Make a list of the final-grade scores in rank order from highest to lowest. Count down from the top the number of cases corresponding to the percentage of students to receive A's. Then, from that point, continue to count down further the number of cases corresponding to the percentage of students to receive B's, etc. This procedure is satisfactory for small numbers of students, say 50 or fewer, but it is a time consuming process when grades have to be assigned to a large number of students.

One method that can be used to estimate the "cut-off" points for the various letter grades is to prepare a frequency distribution such as that shown on page 11. The steps are as follows:

1. If the course consists of separate sections for boys and girls, determine the percentage of boys (to the nearest whole percent) and the percentage of girls.

2. From the sections for girls, draw a spaced sample (every n' th case) equal to the percentage of girls. For example, if the overall percentage of girls in the Course is 64%, draw 64 cases from the sections of girls. If there is a total enrolment of 180 girls select every third case from each section and draw the remaining four cases randomly from all sections. Follow a similar procedure in the case of the boys.

3. Enter the tally-marks on the interval scale as shown in page 11.

4. Since each case represents one percentage point, count down from the top the number of cases equal to the percentage of A's to be given. Repeat the same procedure for the other grades.

5. If in counting down the number of cases to receive a particular letter grade, you find that there are more cases in the bottom interval than you need, you can do one of two things (a) give all students in that interval the grade in question (which will increase the number of students receiving that grade above the number intended) or (b) assume that the tally marks for the interval in question are equally distributed over that interval and estimate the cut-off score for the number of cases you require.

Another method that can be used requires a calculator. First the mean and standard deviations of the distribution of the final standardized scores are determined.

Then the normal curve table is used to determine the cut-off scores that will give the desired percentage of cases in the upper end of the distribution. For further information in this method, consult the Tests and Measurement Office.

Table 2

ELU _____ Semester _____

Course _____ Acad. Yr. _____

Intervals	Tallies	N	
640 +	1	1	
630 - 639	1	1	
620 - 629	1	1	
610 - 619	11	2	A
600 - 609	111	3	
590 - 599	111	3	
580 - 589	1111	4	N=15
570 - 579	111	3	
560 - 569	1111	4	
550 - 559	1111	4	B
540 - 549	111	3	
530 - 539	11111111	8	
520 - 529	111	3	N=25
510 - 519	1111	4	
500 - 509	1111111	7	
490 - 499	11111	5	
480 - 489	111111	6	
470 - 479	1111	4	C
460 - 469	1111	4	
450 - 459	1111	4	
440 - 449	11111	5	
430 - 439	111	3	N=42
420 - 429	1111	4	
410 - 419	11	2	
400 - 409	11	2	D
390 - 399	111	3	
380 - 389	111	3	N=14
370 - 379	11	2	
360 - 369	1	1	F
360 -	1	1	N=4
		100	

A Priori Letter Grade Distribution

A - 15 per cent
B - 25 per cent
C - 40 per cent
D - 15 per cent
F - 5 per cent

Total 100 per cent

Cut-offs for Letter Grades
(from frequency distribution)

A,	580	and above
B,	520	to 579
C,	430	to 519
D	380	to 429
F	379	and below

DJM/st

Tests and Measurement Office
Language Center
Kuwait University

M-25

Staff Assignments in Connection with the Administration of Placement Tests and the Processing of Answer Sheets

A. Roles associated with the administration of the tests.

1. Supervisor of the Test Administration

   The Unit Supervisor usually assumes this job. It entails overall planning, supervision and coordination of the activity.

2. Assistant Supervisor of the Test Administration

   The Unit Test Representative generally functions in this capacity. Duties include, but are not limited to: briefing participating teachers on their duties and responsibilities; distributing test materials to, and receiving them back from, head proctors; observing the various testing rooms before and during the tests to assure satisfactory physical and administrative conditions. NOTE: If a new Unit Test Representative is to begin his or her duties in September and if the UTR for the prior academic year is still on the Unit staff, the prior UTR should assume the aforementioned responsibilities, assisted by the new UTR.

3. Head Proctors (Proctors-in-Charge)

   A Head Proctor is to be designated for each testing room. The head proctor is responsible for receiving from the UTR the test materials for his room, preserving their security during the test, and returning them to the UTR following the test. The head proctor is also responsible for reading oral instructions, for timing the various parts of the test, and for general supervision of the proctors assigned to his or her room.

BEST COPY AVAILABLE

4. Room Proctors

At least one room proctor (in addition to the head proctor) is to be assigned to each testing room for every 25 examinees, or fraction thereof. Proctors are to assist in distributing and collecting test materials, in preventing cribbing, and in assuring that students follow instructions correctly.

5. Floating Proctor

A floating proctor is an Arabic-speaking person whose major task is to answer questions that examinees may have who are unable to comprehend instructions in English. One floating proctor is to service two or more testing rooms, depending on the capacity of the rooms. In making head-proctor and room-proctor assignments, the supervisor should attempt to assign at least one of his or her Arabic-speaking teachers to each testing room, thereby minimizing the need for floating proctors. Units that will require the services of Arabic speakers from outside the unit are to notify Mr. Salah Sabry as soon as possible of their requirements.

B. Roles Associated with the Processing of the Answer Sheets

1. Supervisors of the Test Administration (see A. 1).
2. Assistant supervisor of the Test Administration (see A. 2).
3. Scorers of the multiple-choice parts of the test.
4. Readers of the essay and/or supply-type parts of the test (if applicable).
5. Tabulator.

The tabulator assembles the test-score data in a form that enables the T&M Office to make a quick check on the reasonableness of the cut-off points that have been proposed on the basis of the norming administration and other relevant information. See M-26 of the T&M Office Test Development Manual for instructions concerning the use of the "Placement Test Tally Sheet."

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-26 | مكتب الاختبارات والقياس

Date: June 23, 1980 — التاريخ: _____

Ref: VM/ST — الرقم: _____

MEMO

From: Victor N. Mason

To: Director, Asst. Directors, Supervisors, Coordinators, UTRs

Subject: M-26 of Test Development Manual: Placement Test Tally Sheet

1. Enclosed is a copy of the Placement Test Tally Sheet for the use of tabulators (generally the UTRs) in assembling the score data from the administration of the placement test in a form convenient for checking on the reasonableness of previously suggested cut-off points. These cut-off points are determined primarily---but not exclusively---on the basis of the results of the norming-test administration involving the same test items and sections taken the previous academic year by students enrolled in the same courses into which newly entering students are to be considered for placement.

2. The tally-sheet form consists of six pages for the entering of test-score data. It has been prepared to handle placement tests which go up to a maximum of 160 points. If any future placement test forms are longer in terms of time or actual length, or both, so as to exceed 160 points, one or more pages would have to then be added to accommodate the greater length up to the new point total.

3. The tabulator begins his task by turning to the last page (page 6) of the form, on which the results of students with the lowest possible placement test scores are registered with tally marks in column 2. If no student has a placement test score as low as 10, the tabulator turns to page 5. If, as a hypothetical case, the lowest mark on the test is 23, the tabulator places a mark in column 2. If only one student obtained this score, the numeral "1" is written into column 4 ("FREQ.," for "frequency"). This number (1) is then added to the total of all scores below it (i.e., 0) to give a cumulative frequency of 1, which is written into the corresponding blank in column 5 ("CUM. F."). If some 280 students took the placement test, the percentage of all students with this score (23) and below is then calculated (1/280 = 0.34%) and written into the proper blank in column 6("CUM. %").

4. If the next lowest score is 29, and if 3 students obtained this score, 3 tally marks would be placed in column 2 beside this score. The numeral "3" would then be written into the appropriate blank in column 4. The total of 3 would then be added to the previous total of all scores lower than 29 (i.e., only 1) to give a new sub-total of 4 and written into the blank in column 5. The cumulative percentage of this level would then be calculated (4/280 = 1.4%) and entered into the appropriate blank in column 6.

5. This procedure would then be followed until all 280 students had been accounted for, or 100 per cent of the group taking the placement test. The supervisor and UTR are then prepared to examine the previously recommended cut-off points for the entering class, as suggested by the previous year's norming test results and subsequent report. Determination of these cut-off points is then made in the light of <u>a priori</u> administrative decisions of previous years as to roughly what percentages of students are expected to be placed into each of the various required courses of the unit.

TALLY SHEET

RANGE: 160 - 131

TABULATOR _____

FACULTY _____

TEST FORM _____

DATE _____

TEST SCORE	TALLIES	TEST SCORE	FREQ.	CUM.F.	CUM.%
160		160			
159		159			
158		158			
157		157			
156		156			
155		155			
154		154			
153		153			
152		152			
151		151			
150		150			
149		149			
148		148			
147		147			
146		146			
145		145			
144		144			
143		143			
142		142			
141		141			
140		140			
139		139			
138		138			
137		137			
136		136			
135		135			
134		134			
133		133			
132		132			
131		131			

RANGE: 130 - 101

TABULATOR _____

TALLY SHEET

FACULTY _____

TEST FORM _____

DATE _____

TEST SCORE	TALLIES	TEST SCORE	FREQ.	CUM.F.	CUM.%
130		130			
129		129			
128		128			
127		127			
126		126			
125		125			
124		124			
123		123			
122		122			
121		121			
120		120			
119		119			
118		118			
117		117			
116		116			
115		115			
114		114			
113		113			
112		112			
111		111			
110		110			
109		109			
108		108			
107		107			
106		106			
105		105			
104		104			
103		103			
102		102			
101		101			

TALLY SHEET

RANGE: 100 - 71

TABULATOR _____

FACULTY _____

TEST FORM _____

DATE _____

TEST SCORE	TALLIES	TEST SCORE	FREQ.	CUM.F.	CUM.%
100		100			
99		99			
98		98			
97		97			
96		96			
95		95			
94		94			
93		93			
92		92			
91		91			
90		90			
89		89			
88		88			
87		87			
86		86			
85		85			
84		84			
83		83			
82		82			
81		81			
80		80			
79		79			
78		78			
77		77			
76		76			
75		75			
74		74			
73		73			
72		72			
71		71			

RANGE: 70 - 41

TABULATOR _____

TALLY SHEET

FACULTY _____

TEST FORM _____

DATE _____

TEST SCORE	TALLIES	TEST SCORE	FREQ.	CUM.F.	CUM.%
70		70			
69		69			
68		68			
67		67			
66		66			
65		65			
64		64			
63		63			
62		62			
61		61			
60		60			
59		59			
58		58			
57		57			
56		56			
55		55			
54		54			
53		53			
52		52			
51		51			
50		50			
49		49			
48		48			
47		47			
46		46			
45		45			
44		44			
43		43			
42		42			
41		41			

RANGE: 40 - 11

TABULATOR _____

TALLY SHEET

FACULTY _____

TEST FORM _____

DATE _____

TEST SCORE	TALLIES	TEST SCORE	FREQ.	CUM.F.	CUM.%
40		40			
39		39			
38		38			
37		37			
36		36			
35		35			
34		34			
33		33			
32		32			
31		31			
30		30			
29		29			
28		28			
27		27			
26		26			
25		25			
24		24			
23		23			
22		22			
21		21			
20		20			
19		19			
18		18			
17		17			
16		16			
15		15			
14		14			
13		13			
12		12			
11		11			

RANGE: 10 - 1

TABULATOR _____

TALLY SHEET

FACULTY _____

TEST FORM _____

DATE _____

TEST SCORE	TALLIES	TEST SCORE	FREQ.	CUM.F.	CUM.%
10		10			
9		9			
8		8			
7		7			
6		6			
5		5			
4		4			
3		3			
2		2			
1		1			

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-27 | مكتب الاختبارات والقياس

Date: April 1, 1981 — التاريخ: _____

Ref: T&M Memorandum of November 15, 1980 — الرقم: _____

MEMO

From: Victor W. Mason VM

To: Holders of T&M Manual

Subject: Pre-Test Cover Page: M-27

This memo is a follow-up to that of November 15, 1980, concerning a proposed format for a suitable cover page for copies of new examination forms to be pre-tested within a unit as part of the normal review process for those exams. The memo of November 15, 1980, solicited suggestions for further modifications that staff members might regard as useful or necessary.

The enclosed form may now be considered the final form of the cover page for pre-test copies. Please insert the page into your manual as M-27. The entry in your Table of Contents may be entitled "Pre-Test Cover Page."

PLEASE HANDLE CAREFULLY.

PRE-TEST COVER PAGE

A. Identification Data

1. Name of Course: (e.g.) Science 162
2. Name of Examination: (e.g.) Final Examination
3. Name of Testee: _____
4. Date Distributed: _____
5. Date to be Returned to Test Development Committee: _____
6. Number of Test Booklet: _____
7. Pretest Form: ☐ Preliminary (First)
   ☐ Revised (Second)
   ☐ Other (e.g., Third)

B. Security Instructions

1. Please handle this pre-test form very carefully. Do not leave it lying around your office exposed where students may see it.
2. Please return it to the appropriate Test Development Committee in the same envelope in which it was distributed to you.
3. It is best to receive this form by hand from the appropriate person concerned and also to hand it back to the same person, to avoid later wrangles over tests which were allegedly left somewhere for someone but subsequently disappeared.

C. Procedural Steps

1. Take this test uncritically just as a student would take it, answering all questions as instructed.
2. Put a check mark or question mark beside any line in which you find an obvious error, a source of confusion or some other problem indicating possible weakness in the test, in your opinion.
3. Finish taking the test like a student before addressing possible problems.
4. When you have finished taking the test, turn to the last page of this booklet, which is blank. On this page, please elaborate on any points in the test-form of which you think you would recommend modification to the Test Development Committee.

PLEASE HANDLE CAREFULLY.

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-28 | مكتب الاختبارات والقياس

Date: April 1, 1981 | التاريخ:

Ref: | الرقم:

MEMO

From: Victor W. Mason VWM

To: Holders of T&M Manual

Subject: Addition of M-28 to T&M Manual, M-28: Purpose of the Norming Test at KULC

At the suggestion of Dr. Malcolm, the enclosed memorandum entitled "Purpose of the Norming Test at KULC" is being inserted into the T&M Manual as M-28. Please add this title also to the Table of Contents of your copy of the manual. Thank you.

Purpose of the Norming Test at KULC

1. Kuwait University Language Center requires all students newly enrolled in a college to take the English Placement Test (EPT) of the English Language Unit (ELU) of that college. The main purpose of the EPT is to attain maximum homogeneity in English proficiency of students at each course level in the sequence of required English courses of each ELU.

2. The EPT in each ELU is administered twice each academic year; i.e., at the start of each semester. Each such administration of the EPT is known as the "operational administration" of the test; that given in September is, of course, the more important, since the great majority of new students enroll in a college in the fall following their graduation from high school the previous spring.

3. In order for the Language Center to determine the suitability of its Placement Tests prior to their operational administration, a full-scale pretest of each EPT is done during the academic year preceding the scheduled date of operational administration. For example, EPTs given in September 1980 were pretested in December 1979. The subjects of the pretest in all units were continuing students ---i.e., those already enrolled in the same required English courses that new students would be placed in on the basis of their EPT scores the following academic year.

4. This pretest is known as the Norming Test (NT). The results of "norming" the test provide an indispensable standard by which to compare the English proficiency of incoming students and continuing students. Thus, the NT is extremely important to the staff members of each ELU in providing them a means to achieve a reasonable degree of homogeneity in their EFL classes.

5. As all experienced foreign-language teachers know, the first requirement of a rational program is sufficient similarity of ability levels in a class to make effective instruction possible for all students. In poorly organized classes, a significant number of students will be wasting their time, and will therefore become frustrated---either those who are much better than the majority and will thus quickly feel bored; or those who are much weaker than the majority and will therefore eventually give up; or both. The likelihood of teacher frustration in programs with such classes is also very high.

6. KULC Norming Tests are crucial in the effort to attempt to minimize such frustration on the part of both students and teachers. They are crucial, therefore, in our efforts to promote a foreign-language program of professional quality. They help to do so in three ways:

a. by providing an empirical basis of comparing the English proficiency of incoming students with that of continuing students

b. by acting as a common test for all continuing students already enrolled in an ELU's sequence of required courses

c. by permitting statistical analysis of the items and the parts of each test prior to its operational administration.

Date: June 1980 التاريخ :

Ref : VM/ST الرقم :

Survey:

Tests and Measurement Office

June 1980

A. History of the Office

1. The T&M Office of Kuwait University Language Center was founded by Dr. Donald J. Malcolm in the fall semester of the 1977-78 academic year. Dr. Malcolm was the head of the office during the three full years which he spent with the Language Center, from September 1977 until June 1980.

2. During those three years, he was assisted in the office by several Language Center staff members: Mr. Steven Spurling, Miss Sana'a Al-Naqeeb and Mr. Victor Mason. Mr. Spurling assisted full-time in the spring semesters of 1978 and 1979. Miss Naqeeb was with the office during the spring semester of 1979. And Mr. Mason assisted full-time during the 1979-80 academic year.

3. The T&M Office has worked closely, in the development of tests, with all of the English-language units of the Language Center as well as with the French Language Unit. Nearly all test-development work until now has concerned preparation either of placement examinations by the five ELUs and by the FLU or of final examinations of required courses in the ELUs.

B. Liaison with the Units

4. Liaison between the language units and the T&M Office has been the responsibility of the unit test representatives (UTRs), language teachers in the units who are appointed by their supervisors to coordinate testing activities both within their units and between the T&M Office and their respective units. The considerable amount of time that the job entails has required substantially reduced teaching loads for the UTRs.

C. Preparation of Placement Tests

5. The UTR has the primary responsibility in the unit for over-seeing and coordinating development of new forms of the unit's placement examination for the fall semester of the year, when the majority of new students enter the university. Preparation of the new placement examination begins one full calendar year in advance of its administration, according to the following schedule:

a. early September: preparation of "Test Development Schedule" by T&M Office

b. September-November: Preparation of test

1) writing of items and sections in unit
2) review of items and sections in unit
3) review of items and sections in T&M Office
4) final revisions in unit
5) preparation of two parallel one-hour half-test forms, the "norming test"
6) completion of test-item assembly form (TIAF)

c. late November: printing of norming test in AVA Office

d. December: administration of norming test

e. December-January: computer analysis of norming test results

f. February: analysis of norming test results, and recording of statistical characteristics of items on item cards kept on file in T&M Office; estimating cut-off points on fall placement test

g. March: correction or replacement of faulty items and sections, as determined by statistical analysis; review of corrections by Tests and Measurement Office

h. late March-early April: assembly of the two one-hour tests as the two-hour fall placement test; preparation of NT report

i. late April: printing of fall placement test in AVA Office

j. May: packaging, labeling and storing of placement tests, scoring keys and instructions for administration under secure conditions in the unit

k. May: spring administration of norming test, if necessary (as in ELU Engineering), and combination of results with those from earlier administration for establishment of cut-off points on upcoming fall placement test

l. September: administration of placement test and completion of tally sheet of results to aid in determining cut-off points between classes

6. Following administration of the placement test, the results are computer-analyzed and the T&M Office prepares a report on the results of the test. The UTR records the statistical characteristics of each item for that administration on the item cards.

D. Preparation of Final Examinations

7. The second major area of test-development activity by the UTR concerns preparation of his unit's final examinations for required courses. The UTR is an ex officio member of all test development committees (TDCs) working on those finals, and keeps the T&M Office informed of the progress of the development of those finals.

8. Development of final examinations for required courses in each semester involves the following procedures:

a. preparation by the T&M Office of a "Test Development Schedule" for each unit's final exams

b. submission of these schedules by the UTR to his supervisor and coordinators for their concurrence

c. appointment by each relevant coordinator of a TDC from among instructors teaching the course that semester, with the coordinator as chairman

d. revision of the statement of instructional goals (behavioural objectives) of the course in the event of curricular changes during the given semester

e. preparation of the Final Exam Content Specifications (FECSs), the "blue-print" of the language skills to be assessed, the kinds of test items each section will contain, and the relative weight to be assigned to each

f. discussion between the T&M Office and the UTR of 1) the behavioural objectives and 2) the FECSs for the new finals in the light of a review of past finals for the same courses, with either T&M approval or recommendations for changes

g. following agreement on the FECSs between the T&M Office and the TD Committee, writing and reviewing of items for the new final exam, with the technical assistance of the UTR

h. following assembly of the new test, review of the preliminary finished form by all instructors teaching the respective course and by at least two others not teaching it

i. following all review procedures, submission of the final exam to the unit supervisor for preliminary approval

j. submission, together, of all final exams of a given unit to the T&M Office for review

k. modification of the final by the TD Committee following recommendations by the T&M Office mutually agreed upon

l. submission of final by TD Committee to unit supervisor for final approval

m. submission of final by unit supervisor to T&M Office for final review of test format, answer sheet design, and scoring procedures.

9. The various units have generally arranged to have printing of their final examinations done without the assistance of the T&M Office. Moreover, the statistical analysis of final exams has, thus far, seldom been undertaken, with the notable exception of the final examination for the students of the 098 (intensive) program of the Engineering ELU.

10. It has been arranged, however, that this summer for the first time computer analysis of the Language Center's final exams will be performed at the university's Computer Center on the Adeliyah campus. Once installation of the

new and enlarged Computer Center has been completed on the Khaldiyah campus at the end of the 1980 calendar year, analysis of the Language Center's final examinations at the Computer Center is expected to become a routine procedure.

E. Assistance of the Computer Center

11. Up until calendar year 1980, the T&M Office has obtained the indispensable assistance of the Computer Center in standard item and test analysis. This has been of value primarily in two respects: 1) evaluating achievement of language skills by students, and 2) 2) evaluating the effectiveness of tests themselves in performing that task.

12. The new Computer Center on the Khaldiyah campus will be equipped with an optical scanner which will "read" answer sheets up to a maximum capacity of 25,000 sheets per hour. This device will relieve the Language Center's teachers entirely of the need to score objective tests such as the units' placement tests, as well as large parts of the finals of many courses.

13. Answer sheets will become standardized, to utilize only a limited number of formats. Early in the fall semester of the 1980-81 academic year, the T&M Office is to consult closely with the personnel of the Computer Center to work out the specific formats for answer sheet design desired by the ELUs of the Lanugage Center. Once these formats have been settled upon, printing orders will be sent from the Computer Center to the printers in England.

14. As part of the expansion of Computer Center services, terminals will be installed in many offices of Kuwait University. One such terminal has been requested by the Language Center to facilitate the storage and convenient retrieval of data on students relevant to Language Center needs. A terminal will greatly aid the Language Center

in several ways. For example, at registration time an official can see at a glance all language courses taken by a given student. The student is thus much less likely to register improperly for a Language Center course, such as to attempt to bypass prerequisite courses, as has happened not infrequently in the past.

15. The terminal will be invaluable also for permitting the Language Center to have ready access to each student's entire record of course marks and his scores on all major "Language Center-level" tests, such as placement tests, mid-term tests, "unit" or common tests, norming tests and final examinations. The convenience and usefulness of this new system for record-keeping, for data retrieval and for evaluation purposes can hardly be overestimated.

F. The Test Development Manual

16. During the years 1977-80, Dr. Malcolm prepared a large number of memoranda to guide the units in the development of professional-quality language tests. These many memos have been assembled to constitute the T&M Office's Test Development Manual.

17. Some of these memoranda deal with the duties of the UTRs, coordinators and supervisors in the development of the placement tests and final examinations. Others deal explicitly with the procedures to be followed in the writing and reviewing of specific types of test items and sections. The guidelines deal with both objective-type and supply-type testing formats.

18. The manual provides guidance in the specification of curricular goals of the Language Center's courses. It provides assistance to the staff in interpreting the results of computer analysis of major examinations. It gives a detailed explanation of how to fill out forms for the processing and summarizing of essential test data. It explains how to compute standard scores, to minimize the effect of teacher subjectivity in the assigning of grades to students enrolled in courses with multiple

sections whose teachers may vary widely in their grading practices.

G. Record-Keeping Systems

19. The T&M Office has several types of record-keeping systems. The most important ones are the filing system, the card catalog and the computer print-out records.

20. The files maintain a complete set of old ELU test forms for the past three years, both for placement tests and for the final examinations of required courses. Included with the finished test forms (both base and scrambled forms) are answer sheets, answer keys, scoring transparencies and rationales for the scoring of supply-type material.

21. The files for placement tests also contain reports of the results of those administrations. Separate files contain norming test forms and reports on the results of those test administrations, as well.

22. The files for course-final examinations are accompanied by additional files stating curricular objectives and the Final Exam Content Specifications for each course. These are kept updated each semester.

23. A large part of the filing system is given over to Language Center, unit and T&M Office administration. Several sections of files are primarily devoted to research and evaluation reports prepared by Dr. Malcolm. Some of these reports concern programs outside the Language Center, such as studies done for the College of Graduate Studies on the basis of experimental administrations of the TOEFL. Much space in the filing system is unfortunately taken up by enormous numbers of answer sheets.

24. The card catalog is maintained to provide an item bank for the preparation of new placement test forms. All placement tests of the Language Center are in a multiple-choice format, and the items---primarily to test grammar,

vocabulary and reading comprehension--- are maintained each on a separate card. The card catalog system of placement test items was meant to serve as a model for the development and storage of final examination items as well, but until now this ambition has been unfulfilled.

25. All computer analysis of Language Center test results has been performed on either placement tests or norming tests. The Engineering 098 final multiple-choice test used to be analyzed only because it was the unit's norming test for 098 students as well. These results are maintained on computer printouts bound into book form. The need for greater efficiency in binding these printouts will increase now that many course-final examinations are being computer-analyzed for the first time during the summer of 1980.

H. Continuing Problems

26. Probably the principal difficulty for the T&M Office semester after semester is a too-limited understanding by many members of the Language Center's teaching staff of what the office is attempting to do. Characteristic of the problem are statements from course coordinators that suggestions from the T&M Office result in a lot of extra work for the units---the implication being that much of that work may be unnecessary. These complaints generally increase toward the end of a semester when pressures on TD committees to finish their final exams on schedule are greatest.

27. The subject of development of final examinations for required courses is without doubt the source of more frustration to the T&M Office than any other. The reasons for dissatisfaction are several, all interrelated. First and foremost is the problem that many TD committees simply do not adhere to the test-developement schedules prepared at the start of a semester by the T&M Office.

28. This failure to adhere to the test-development schedules not infrequently results in final exams being submitted late to the T&M Office for its review, when recommendations for change can no longer be implemented. Experience has demonstrated repeatedly that the more tardy the completion of a final exam, the more likely that test will suffer from major structural flaws. Such weaknesses are the consequence of neglect of probably the most important part of the test-development process: a sufficiently thorough and adequate review, involving pre-testing, thorough criticism, rewriting and further review and revision.

29. One of the principal reasons for weak tests is the very uneven abilities of EFL teachers in preparing acceptable tests: many are very good at writing test items and sections but many others, through lack of training or experience, or both, are less so. The recommended system of preparing final examinations by committee over several months is intended to try to neutralize any test-writing deficiencies of Language Center staff. Guidelines to the writing and reviewing of tests are laid out explicitly in memos of the TD Manual written by Dr. Malcolm to assist the staff. From the appearance of many final examinations, however, it appears as if those guidelines are too infrequently consulted.

30. It is widely known that the preparation of good multiple choice items is difficult and time-consuming. Language Center finals resorting to this format, especially to test reading comprehension and grammar, often are insufficiently reviewed to remove or improve obviously weak items.

31. Many final examinations in the past have not been comprehensive enough. It has been possible in some courses for all students, even the weakest, regularly to complete their finals in less than two hours, for courses that have met every day of the semester.

32. It is not difficult each semester to find a final examination for a higher course which is obviously less challenging than the final examination for a lower course

in the same unit. This is despite the fact that the lower course is supposed to be a prerequisite for the higher.

33. Some TD Committees use the same test sections in the final examinations of their courses two or more semesters in succession. For testing reading comprehension, listening comprehension and writing, this practice can clearly compromise test security. Such frequently used sections will be discussed by students and will give an unfair advantage to students in following semesters knowing what to expect.

34. In the Language Center, the T&M Office has the ultimate responsibility for assuring quality control of major "Language Center-level" examinations. The office cannot, however, take responsibility for final examinations that it receives too late to review, prior to test administration.

35. The FE development process begins at the start of a new semester, so it does not seem unreasonable to ask that the preliminary finished form be submitted within 2½ to 3 months, as specified by the TD schedule and agreed to by the TD Committee.

36. The limited resources of the T&M Office necessitate a schedule for review and revision of FEs that is early and long enough to make the orderly review of the final exams for all required courses possible. This process should allow a minimum of two weeks when the T&M Office can give its undivided attention, during office hours, to the completion of all reviews of finals, the preparation of summary reports (when necessary or desirable) and incidental discussions with concerned UTRs and/or TD Committees.

VM/ST

A Brief Biography of Dr. Donald J. Malcolm

Dr. Donald J. Malcolm received his M.A. from the University of Rochester in 1947 and then taught courses there for a year in statistics and psychology for teachers. After that, he attended Harvard University, where he was a teaching fellow and research associate at the Center for Field Studies of the Harvard School of Education, obtaining his Ed.D. in 1951.

Following graduation, Dr. Malcolm worked in program evaluation for the Human Resources Research Institute of the U.S. Air Force at Maxwell Field, Alabama, from 1953 to 1959. He joined the Educational Testing Service (ETS), at Princeton, N.J., in 1960 and remained there nearly two decades.

While with ETS, Dr. Malcolm held a series of highly responsible positions. He became assistant director for test development, responsible for (a) the College Board Achievement Test and (b) the Advanced Placement Test. From 1967 to 1969, he was advisor to the Board of Intermediate and Secondary Education in Bangladesh, where he was Program Officer for the Ford Foundation.

In 1969, Dr. Malcolm took over as the third director of the world-wide TOEFL program, following David Harris and Leslie Palmer. He held that position until 1977, longer than any other TOEFL director thus far.

From 1977 till 1980 Dr. Malcolm was head of the Tests and Measurement Office of Kuwait University Language Center. He left in June 1980 and it is hoped he will return in the future as a consultant to the Language Center.

VM/st

by Victor W. Mason
Tests & Measurement Office
Language Center
Kuwait University

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office | M-30 | مكتب الاختبارات والقياس

Date: May 26, 1981 التاريخ:

Ref: Memorandum of 11 May 1980 to Unit Supervisors الرقم:

MEMO

From: Victor W. Mason

To: All Holders of TD Manual

Subject: Security Systems for 1) Tests Based upon the Multiple-Choice Format and 2) Items Banks

1. The enclosed new memorandum to be inserted into the Test Development Manual as M-30 is concerned with the following policies, practices and procedures for maintaining and assuring security for KULC examinations heavily dependent on the use of the Multiple-Choice format:

A. security for item banks

1. the importance of student ignorance about the existence of item banks
2. the numbering of test booklets and answer sheets
3. student requests to review their tests
4. the destruction of old test booklets
5. storage systems for item cards

B. security procedures for the administration of major examinations

1. the use of alternate test forms
2. the use of test-designation and form-designation codes
3. test-distribution patterns to foil student cheating

4. seating of students by teachers, and

5. realistic timing of test parts, whether formally or informally.

2. It cannot be emphasized too strongly that the Language Center's first line of defense against widespread cheating on exams and compromise of test security systems remains student ignorance as to what those systems are. Given the emphasis on tests and grades at Kuwait University, it would be naive to assume that test results and grading considerations, unfortunately, are not the primary concern of a very large number of students enrolled in university courses that are within the Language Center's jurisdiction. For those determined to use underhanded means of trying either simply to pass a course or to obtain a grade they do not deserve, knowledge concerning the heart of our security systems may be valuable indeed.

3. It is my strong request, therefore, that to prevent large-scale cheating in an examination room henceforth, KULC staff be extremely reticent with students on any subject concerning security systems for maintaining the integrity of future examinations depending extensively on the MC format. Staff should never discuss these subjects with the students, should avoid telling them which teachers of a course are most responsible for development of course common exams (e.g., members of the Test Development Committees), and should avoid telling them anything about any office or location where tests may be developed, stored, or reproduced; i.e., as regards their own units, the AVA offices and the Tests and Measurement Office.

4. The students would in any case be curious about these subjects. Teachers should write down the names of students going out of their way to ask them improper questions relating directly to test security systems and procedures, and should give those names to their unit supervisors. It goes without saying that staff members raising such subjects with students would not be acting in a responsible and professional manner.

5. It is hoped that these matters can be clarified for all staff members of the units. It is also highly desirable that awareness of KULC policies and practices of test security henceforth be made an integral part of the orientation procedures for new KULC staff members.

KUWAIT UNIVERSITY
LANGUAGE CENTRE

جامعة الكويت
مركز اللغات

Tests and Measurement Office

مكتب الاختبارات والقياس

Date : May 26, 1981 التاريخ :

Ref : VM/ST الرقم :

MEMO

From Victor W. Mason VWM

To Holders of TD Manual

Subject Security Systems for Tests Based upon the Multiple-Choice Format and for Item Cards

A. Introduction

1. The multiple-choice (MC) testing format is one of the most useful developed in recent decades but also one whose security is among the easiest to compromise. The fact that MC items are also among the most difficult, time-consuming and expensive to prepare makes close attention to security systems for protecting MC tests and item banks essential.

2. The extra efforts required to protect the integrity of the system are repaid many times over by the time saved for teachers in subsequent test development work and by the demonstrable improvement in test quality over time. The data recorded on the basis of test and item analysis are also one indispensable basis for undertaking meaningful program evaluation and revision.

3. In the Language Center, an item bank now exists for all Norming Tests and Placement Tests used by the HIIs. Therefore, stringent security procedures are followed before, during and after each such administration. Now as item banks are developed for the final examinations for many KULC required courses (and even for some major progress tests: mid-term exams, "unit exams," etc.), it will be necessary for similar procedures

to be instituted by the Test Development Committees (TDCs) for these courses.

B. Security for Item Banks

4. It is, therefore, not too soon to specify the indispensable security precautions that must be taken, to avoid many hundreds of man-hours of teachers' test-development time going for naught. It should be apparent, for example, that the disappearance of a single test booklet, before or after an exam, can wipe out a large part of a TDC's item bank. Given the determination of many of our weaker students to pass their finals, even if by dubious means, KULC staff members must do their utmost to safeguard test security at all times.

1. Student Ignorance about Item Banks

5. The first security requirement for safeguarding a course item bank is student ignorance about its existence. A test booklet that disappears following administration of an exam obviously becomes much more valuable to a student and to his friends if they can be sure that the identical items will reappear in the future. Just one teacher thoughtlessly discussing the existence of such a system with students compromises the security of tests not only for his or her course; students would be quick to assume that a similar system exists also for other courses. Students transferring from one college to another could quickly transmit this useful information and put a premium on the assembly of KULC test files, much as test files are accumulated at universities abroad (e.g., by fraternities in the United States). Thus, discussion about the system by a teacher with his or her students would be very irresponsible and unprofessional.

2. Numbering of Test Booklets and Answer Sheets

6. One essential procedure for guaranteeing security for an item bank before and after test administration is to number all test booklets and attached answer sheets. This not only greatly simplifies the packaging of new test forms and their distribution and collection in the examination room; it makes it possible to trace any missing booklet and/or

answer sheet, especially in double-checking tests when handed in, in the examination room, and in triple-checking them later.

7. Teachers responsible for administering a test can more easily check for accuracy before the test administration begins, to see that the totals and numbers written on the outside of the test package correspond to its contents. Inside the testing room, they can distribute their tests in serial order and collect them later in a similar manner. Furthermore, the numbers on the test booklets and answer sheets greatly facilitate the efficient double-checking of totals at the end of the exam, when students, having been instructed to remain in their seats, begin talking and when confusion may thus set in. These procedures are already followed in the case of all KULC Placement Tests and Norming Tests.

8. The fact that both the test booklets and the answer sheets are numbered has several distinct advantages. Students who absent-mindedly (i.e., innocently) carry off a test booklet (it does happen!) are highly unlikely to forget to turn in their answer sheets; the latter have their names on them so they may be rapidly identified and the booklets recovered. For those rarer cases of students---aware they are failing badly---who might try to conceal both booklet and answer sheet in their clothing (to prepare for a later day), numbering is not enough; a student attendance roll which all students sign (ideally, in the presence of their own teachers) is also necessary. Simply counting the total number of students is not enough, especially when weak students try sometimes to have friends take their tests for them.

9. One necessary procedure for reducing the likelihood of loss of a test booklet after test administration is to tell students before the test that they may take nothing to their seats except writing implements. Books, purses and other such personal belongings should never be permitted to be carried to their chairs.

3. Students' Requests to Review Their Tests

10. A major security risk to any test system occurs when students are permitted to review their tests following administration and correction. The risk is two-fold: one, the physical handling of the booklet, which may disappear; and two, the opportunity for students to study closely the contents of the booklet, even when they hand back the test form to the teacher.

11. At KULC, this risk is less in the case of final examinations, since these are never returned to the students and may be seen only by special request in the office, and under the supervision, of their teachers. The risk is far greater in the case of course non-final common progress tests (midterms, "unit tests," etc.), since it is a reasonable part of teaching procedure for these to be brought into the classroom and distributed for the students' inspection. Test contents may then be fixed much more firmly in mind, especially the vocabulary items, reading passages and listening passages.

14. Moreover, if an item bank is to be prepared for such non-final common progress tests, the teachers must again collect all of the test booklets and answer sheets and count them both in class before the students leave the classroom and later in their offices. This task is complicated by the fact that, when students from different sections are mixed randomly for the specially scheduled administration of common exams, students of the same class will not have test booklets numbered in sequence. It is also much more difficult during ordinary classes to prevent students' personal belongings from becoming mixed up with the tests themselves, particularly amidst the commotion reigning at the end of a typical class period.

15. For these reasons, item banks for course final examinations must be kept separate from those for all other course common tests. They cannot be mixed. Secondly, the ease with which students remember the contents and results of pre-final common exams means that the items and sections of these tests can be safely re-used far less often than can those of final exams. In the case of a final exam, use of a vocabulary item or a reading or listening passage (or writing task) should probably not be repeated more than once every three or four semesters (i.e., about once every two years). In the case of pre-final common exams, a safe standard would probably be to extend these time restrictions by 50 to 100 per cent (i.e., once every three or four years).

4. Destruction of Old Test Booklets

16. Since every old test booklet developed on the basis of an item bank becomes a potential security risk, immediate destruction of tests becomes imperative. It is inconceivable that they would simply be left lying about for days, weeks and even months at a time. All old test booklets except for those needed as file copies for record-keeping or needed for special make-up test administrations should be destroyed.

17. In the 1981-82 academic year, the KULC AVA Office will have a powerful new shredder installed for this purpose. This machine will be an exact duplicate of that now in the T&M Office, which is available for the destruction of all unit placement tests and norming tests.

5. Storage of Item Cards

18. Every ELU will need a special cabinet for storing its item cards, similar to the one in the T&M Office which holds KULC Placement Test items. Coordinators with temporary security problems for their cards may keep them at home until their unit's own cabinet has been acquired.

C. Security Procedures during the Administration of Tests

19. Security policy and procedures in effect during test administration rely most for their success upon the same basic principle emphasized in the above discussion on item banks: student ignorance of their existence. The procedures and practices employed in the case of major KULC examinations depending heavily upon the multiple-choice format are primarily the following three:

a. the use of alternate test forms, such that multiple-choice items of the base ("A") form are scrambled in the alternate form(s)

b. the use of test designation codes---familiar to the teachers but not the students---distinguishing between the forms, and

c. distribution of test booklets such that no student sits adjacent to another with the same test form.

In addition, the T&M Office recommends:

d. seating of students such that friends may not have the opportunity to assist each other, even clandestinely, during an exam, and

e. realistic timing of test parts, so that students prepared for the exam should be able to complete all all parts and sections.

20. Security during the administration of MC tests is greatly enhanced by having alternate forms of the same examination; i.e., a base form and one or more "scrambled" versions (dubbed "A," "B," "C," etc.). Students cannot cheat from their friends 1) if there are different test forms being used and 2) if they do not know who has the same test form as they do. Students should never be told about the nature of the differences in their tests or about the way in which those differences may be identified.

21. This is one of the principal reasons behind the use of test designation codes at the top of the cover page of a test booklet. The code including the test-form designation (A, Q, X, etc.), is intended to be meaningless to the students. (The existence of such a security code will be pointless if the students are to have the code conveniently broken for them by being told how it works.)

22. Therefore, for course common examinations, including final examinations, these letters are best not included with test page numbers throughout the test booklet. Page numbering that combines the form code and page number has been used in the past for convenience during the reproduction and collating of test booklets; it has no security value. In the future, it would be best that entirely different symbols (preferably unfamiliar to the students---e.g., letters of a foreign alphabet) be placed in an inconspicuous page corner to distinguish between test forms. The same precautions should

be taken with future administrations of the Norming Test; and even with the Placement Test, since some incoming students seem aware of our alternate-form system for MC tests, and therefore have been known to object if not allowed to sit in a seat of their choice.

2. Distribution of Test Booklets

23. The pattern of distribution of test booklets for major KULC examinations employing the alternate test-form system has followed the "diamond" configuration. That is, no student ever sits adjacent to another student with the same form. This arrangement should prevent pairs of neighboring students from assisting each other on these tests, since item numbers for the same tasks do not correspond.

24. However, it sometimes happens that students learn of these arrangements for test-booklet distribution and for test codes. When they do, there is then the serious danger of mass compromise of test security on the multiple-choice parts of an exam. This security breach comes about when---as sometimes happens on KULC final exams---a large group of friends aware of this system arrange to seat themselves such that a student good in English sits in the front of the testing room while his or her weaker friends sit in back in chairs with the same test form. They then work together in tandem through entire test sections or parts, with the one(s) in front giving surreptitious pre-arranged signals to the watchful host in back.

25. Such large-scale cheating is easiest to achieve in MC sections with discrete-point tasks, especially grammar and vocabulary. More integrative tasks like reading are more difficult to cheat on because students work at much different rates and might not undertake reading passages in the same order. "Inherently timed" tasks like listening comprehension and dictation almost by definition prevent such cheating unless students flagrantly talk or whisper to each other.

26. In the event of students attempting to circumvent the aforementioned security systems for MC tests, it becomes necessary to resort to more elaborate security measures. For example, a unit must perhaps prepare for three or more

alternate forms---A, B and C---rather than two. Equally important, students then should not witness the system of distribution to see who has the same test form. It has occasionally been the practice in some ELUs, for example, to distribute all test forms face down in advance and then to seat the students, telling them not to touch their tests booklets until instructed to do so.

3. Seating of Students

27. An important security precaution for all major examnations vulnerable to compromise is the practice of not permitting students to sit where they wish. Students should enter the testing room one at a time and be seated by the proctors. Ideally, teachers who know the students should be their proctors if at all possible, so that friends can not sit near each other. In particular, the best students in each group should be directed to the back of each testing room, where there will be no pressure on them to help their less able classmates.

4. Timing of Major Tests

28. One of the best ways of preventing students from cheating on major examinations, especially on finals, is to time all parts fairly but realistically. Efficiency of language usage in all the language modes is a basic criterion of proficiency and timing of tests simply reflects this fundamental consideration.

29. A test written to help all students finish everything is not being directed at the average majority; it is being prepared at the lowest level of proficiency: those students most in danger of failing a course. In the event of such a test, the great majority of students find it so short, and usually easy, that they can leave the examination room prior to the end of the scheduled test time.

30. Such a test usually leads to a serious imbalance of A and B grades vis-a-vis Cs, Ds and Fs. This is often so because, being too short and easy, it gives the great majority of students of average ability, at that level, too much time to dawdle and to carry out pre-arranged "mutual assistance strategies."

It is in no way challenging for the best students, who do not have to cheat anyway. Under such circumstances, mediocre students can qualify for exam scores and final grades they do not deserve, thus contributing to "grade inflation" and a consequent general lowering of academic standards.

31. Timing can be informal (e.g., advisory) and can be communicated both in the test booklet and on the blackboard during the exam. Or it can be rigorously enforced for each part of a test, so that students are prohibited from looking at any test part other than that which they are instructed to be doing at a given time. In the latter case, test parts must be printed on paper of a different color, so that teachers can see at a glance whether any students are looking at a part of the exam which they are, at that moment, not authorized to be doing.

32. The claim that, during a course common achievement exam, especially a final, students should have a right to be able to review previously completed work is a strong and valid one. The needs of both students and testers can be accommodated (in the case of formal timing) if students are informed in advance that, at the end, they will be allowed a final, say, 10 or 15 minutes (maximum) to quickly skim all parts of their tests to review their answers for errors. They should, of course, be permitted to write at this time, finishing up whatever they wish. It is important that the provision of this formal review time not be abused by giving students more time than is needed by <u>those already finished</u> with all parts to skim their answers. This practice is reasonable, however, only if a test 1) has been formally timed, 2) is comprehensive, 3) is truly challenging throughout for the majority of students, and 4) is long enough both a) to keep all or virtually all students in the test room until the end of the scheduled test period and b) to keep the average majority of students busy writing virtually until the final minutes of the test time allotted.

VM/ST

KUWAIT UNIVERSITY PRESS